

Digitized by the Internet Archive
in 2023 with funding from
University of Toronto

DIRECTIONS

THE FINAL
REPORT OF
THE ROYAL
COMMISSION
ON NATIONAL
PASSENGER
TRANSPORTATION

Volume 3



CAI
Z1
-1989
T611
v.3

DIRECTIONS

THE FINAL

REPORT OF

THE ROYAL

COMMISSION

ON NATIONAL

PASSENGER

TRANSPORTATION



Volume 3

The opinions expressed in Volumes 3 and 4 are those of the authors of the individual studies and do not necessarily reflect the views of the Royal Commission.

We gratefully acknowledge permission granted to reprint previously published and unpublished material by the following: Academic Press, Inc.; American Telephone and Telegraph Company; Pierre Berton; The Brookings Institution; Charles River Associates Incorporated; Federal Reserve Bank of Boston; Federal Reserve Bank of Chicago; Federal Reserve Bank of Cleveland; George W. Hilton; The Institute of Economic Affairs; Institute for Research on Public Policy; John Deutsch Institute, Queen's University; Alfred E. Kahn; Frank S. Koppelman; Douglass Lee; Oceana Publications, Inc.; Organisation for Economic Co-operation and Development; Oxford Review of Economic Policy; Oxford University Press; Pergamon Press Ltd.; University of Toronto Press; Transportation Association of Canada; Transportation Research Board, National Research Council; George W. Wilson; Clifford Winston.

CANADIAN CATALOGUING IN PUBLICATION DATA

Canada. Royal Commission on National Passenger Transportation

Directions: the final report of the Royal Commission on National Passenger Transportation.

Issued also in French under title: Directions.

To be complete in 4 v.

Chairman: Louis D. Hyndman

ISBN 0-660-14092-6 (set of 4 v.)

ISBN 0-660-14545-6 (v. 1 and 2)

DSS cat. no. Z1-1989/1-1992E (set of 4 v.)

DSS cat. no. Z1-1989/1-1992-1-2E (v. 1 and 2)

1. Transportation — Canada — Passenger traffic. 2. Transportation — Passenger traffic — Government policy — Canada. 3. Transportation — Passenger traffic — Environmental aspects — Canada. 4. Carriers — Government policy — Canada. I. Hyndman, Louis D. II. Title. III. Title: the final report of the Royal Commission on National Passenger Transportation.

CONTENTS

Page

VOLUME 3

Preface	xiii
 1 Icons and Albatrosses:	
Passenger Transportation as Policy and Symbol in Canada	1
D.R. Oworm	
1. Introduction	1
2. The Era Before Railways: To 1850	5
3. Land Travel During the Great Railway Era: 1849–1920	11
The Projects of Confederation: The Rise of	
the Railway as National Symbol	21
Passenger Traffic at the Turn of the Century	33
The Final Phase of Railway Development: 1897–1917	41
4. The “Mature” Railway System and the Rise of the	
Car Culture: 1900–1945	46
The Rise of the Car Culture and its Consequences: 1900–1929	51
The Great Depression	73
World War II	81
5. The Competitive Era and the Rise of the Passenger	
Policy Issue: 1945–1967	83
The Rise of Competition in the Post-War Era	85
Cars and Roads in the Post-War Era	88
The Highway Crisis and the Federal Presence: 1949–1967	99
The Airplane	105
The Politicization of the Railway Passenger Issue	113
6. The Modern Era: 1967–1990	123
Subsidies and Cutbacks	126
7. Conclusion: Concerns and Myths in Contemporary Canada	139
 Endnotes	145



2 U.S. Intercity Passenger Transportation Policy, 1930–1991:	
An Interpretive Essay	165
George W. Wilson	
1. Introduction	165
The Submergence of Efficiency Criteria	168
The Distinction Between Passenger and Freight	
Transportation	172
The Role of Economists	177
2. The Period Before 1930	180
Infrastructure Policy	180
Regulatory Policies	187
The Policy of Modal Separation	194
3. The Great Depression	197
Transportation Infrastructure	204
Transportation Policy	207
Regulatory Expansion	209
4. World War II to the <i>Transportation Act of 1958</i>	211
ICC Policy	213
The Weeks Report	217
The Hosmer Report	220
The Situation in Air and Highway Transport	223
Conclusion	228
5. Evolution of Pro-Competitive Policies: 1960–1982	229
The Doyle Report	231
President Kennedy's Message on Transportation	235
After President Kennedy	239
The Economists' Role in Transportation Reform	240
Airline Deregulation	251
Bus Regulatory Reform	253
The Creation of Amtrak: A Deviation from	
Pro-Competitive Policy?	256
Two Studies Regarding the U.S. Transportation System	
in the Late 1970s	261
6. A Decade of Experience Under Reform	263
The Airlines After Deregulation	264
Amtrak in the 1980s	268
The Latest National Policy Report	272

The Unfinished Business of Regulatory Reform	274
Beyond Regulatory Reform	275
Efficient Infrastructure Pricing: Highway User-Charges	277
Establishing Linkages	280
7. Lessons from the Past: Relevance to Canada	282
Why Canada Was First	283
Canadian-U.S. Policy Differences	285
Lessons from the U.S. Experience	291
The Message for Passenger Transportation	293
Sources of Competition in Intercity Passenger Service	295
The Issue of Subsidy	296
Other Messages and Lessons	297
Endnotes	303
 3 The Role of Equity Considerations in the Provision and Pricing of Passenger Transportation Services	 313
Robin Boardway	
I. Introduction: Equity as an Objective of Government Policy	313
1. The Objectives of Government Economic Policy	314
2. The Role of Government in a Market Economy	321
II. Instruments for Achieving Equity in the Allocation of Resources	326
1. Taxes	326
2. Transfers	329
3. Social Programs	331
4. Federal-Provincial Issues	332
5. In-Kind Transfers	335
6. Regulation	336
III. Application to Passenger Transport Services	337
1. General Principles	337
2. Alternative Ways of Addressing Equity Issues in Passenger Transportation	339
3. Subsidizing Passenger Transport as an Instrument of Income Redistribution	340
4. Subsidizing Passenger Transport as a Decreasing Cost Industry	343
5. Passenger Transport Services as Assistance to Persons with Characteristics Other Than Income	343



IV. Summary Remarks	350
Appendix	352
Endnotes	358
References	360
4 Transportation and Economic Development:	
A Survey of the Literature	365
David W. Slater	
Overview	365
The Challenge	366
The Literature That Provides Evidence on	
Transportation and Economic Development	368
Themes in the Literature	369
Geographic Distribution and Movement of People	370
Contentious Analytical and Policy Views	371
The Survey	
Part A — Transportation and Economic Development	372
The Statistical Record of Economic Development	372
Theories of Economic Development	376
Theory of International and Inter-Regional	
Trade, Industrial Policy and Transportation	379
Theories of Public Goods and Public Choice, and	
Impressions Regarding Application to	
Transportation: Issues of Externalities	382
Macro Views on Transportation and Economic Development	387
Disaggregated Macro Models of Infrastructure,	
Transportation and National Economic Development	388
Inventories of Deficiencies of Infrastructure	390
The Literature on Transportation and Economic Development	394
Can Generalizations be Made About Over- and	
Under-Investment in Transportation after Examining	
the Project Evaluation Literature?	402
Regional Infrastructure (Particularly Transportation) and	
Economic Development	403
Economic Development and Intercity Passenger Transport	408

Efficient Transportation Infrastructure Investment and Pricing Policy	412
Trends and Prospects for Transportation Needs	415
Some Tentative Conclusions on Transportation and Economic Development, with Special Emphasis on Passenger Transportation	419
Part B — Transportation and Regional Economic Development (With Special Emphasis on Passenger Transportation)	423
Questions and Concepts	423
Trends of Regional Economic Development in Canada and the United States	424
Consider Canada First	424
Have Transportation Elements Been Major Forces Shaping the Regional and Inter-Regional Structure of the Canadian and United States Economies?	427
Literature to be Surveyed	428
Themes in the Literature	428
Some Details for Various Kinds of Regions	429
Transportation and Economic Development in the Hinterland and Between the Hinterland and Export Markets	432
Conclusions Regarding Transportation and Economic Development	436
Part C — Transportation and Tourism	438
Introduction	438
Transportation Supply	440
Importance of Tourism	440
Modes of Touring	441
Elasticities of Demand	442
Transportation Responses to Tourist Demand and Supply	442
Endnotes	443
References	444
5 Subsidies in Canadian Passenger Transportation	463
Trevor D. Heaver	
I. Introduction	463
Objectives of the Paper	464
Outline of the Paper	464

II. Characteristics of Subsidies	464
Definition of "Subsidy"	465
Types of Subsidy	466
Purposes for Subsidies	472
The Economic Effects of Subsidies	478
Subsidy Management	482
III. The Subsidy Experience	488
Highway Transportation	489
Rail Services	493
Air Services	493
Marine Transportation	494
IV. Assessment	495
Modal Services and Subsidy Types	495
Comparison of Subsidy Practices and Principles	499
Endnotes	502
6 Transportation Infrastructure Policy: Pricing, Investment and Cost Recovery	503
David Gillen and Tae Hoon Oum	
1. Introduction	503
2. Infrastructure Pricing: Theory, Issues and Applications	506
2.1 The Role of Prices and Welfare	506
2.2 Alternative Pricing Methods	508
2.3 Non-Market Mechanisms	517
2.4 Potential Difficulties of Implementation	519
2.5 Applications of Pricing Principles	524
2.6 Summary	531
3. Carrier and Infrastructure Cost Structures:	
A Literature Summary	534
3.0 Introduction	534
3.1 Carrier Costs	537
3.2 Infrastructure Costs	542
3.3 Summary of the Cost Structure for Carriers and Infrastructure	545

4. A Framework for the Analysis of Optimal User-Charges and Cost Recovery	547
4.1 General Framework for User-Charges and Cost Recovery	547
4.2 Pricing and Investment in a Network	549
4.3 Second-Best Pricing and Second-Best Investment	553
4.4 User-Charges and Cost Recovery	555
4.5 Road User-Charges	562
5. Alternative Methods of Financing Road Infrastructure	568
5.1 Introduction	568
5.2 Three Categories of Current User-Payments	569
5.3 Evaluation of the Alternative Financing Tools	570
5.4 Summary of Alternative Road Financing Methods	576
6. Summary and Conclusions	582
6.1 Purpose and Organization of the Study	582
6.2 Infrastructure Pricing Principles	584
6.3 Carrier and Infrastructure Costs	587
6.4 Implementation of User-Charges	588
6.5 Alternative Sources of Financing Road Infrastructure	590
Endnotes	591
References	594
7 Competition Policy and Canadian Passenger Transportation	613
John Blakney	
I. Introduction	613
II. <i>Competition Act</i>	615
A. Overview	615
B. Scope of Coverage	620
C. Enforcement Powers and Practices	624
III. <i>Competition Act</i> — Principal Jurisprudence	626
A. Introduction	626
B. Predatory Pricing and Price Discrimination (Paragraph 50(1)(c) of the <i>Competition Act</i>)	628
C. Conspiracy	637
D. Mergers	644
E. Abuse of Dominant Position	666

IV. <i>Competition Act</i> Compared with <i>National Transportation Act, 1987</i> and Related Legislation	679
A. Introduction —Historical Perspective	679
B. <i>National Transportation Act, 1987</i> and Related Legislation	681
C. Jurisdictional Relationships	684
D. Institutional/Administrative Differences	693
V. Transportation Industry Developments: Appropriate Trade Practices Framework	702
A. Introduction	702
B. Industry Developments	703
VI. Concluding Remarks	716
Charts	719
Endnotes	722

8 Controlling Market Power in Weakly Contestable

Canadian Airline Markets 723

Keith Acheson and Don McFetridge

1. Introduction	723
2. The Problem	725
3. The Approach	725
4. Current Access, Rate and Fare Controls	726
Southern Canada	726
Northern Canada	727
Application of Provisions	728
5. Pricing and Quality Experience	730
Discounting	730
Frequent Flyer Programs	735
Travel Agents and Prices	735
Computer Reservation Systems	736
Overall Canadian Price Performance	736
American Pricing Experience	739
Prices and Number of Carriers	740
Network Reconfigurations, Affiliates and Independents	740



Scale and Network Economies	742
Congestion and Access to Airports	745
Flight Quality	748
Charters in the New System	749
Scheduled Flight Frequency	750
Carrier Competition	750
Profits	753
6. The Number of Carriers and the Exercise of Market Power	757
7. Contestability	758
8. Institutional Defences Against Market Power	767
Further Deregulation	767
Residual Rate or Fare Regulation	768
9. Implications of Northern Routes and Other Mode Regulation for Oversight of Fares on Monopoly Airline Routes in Canada	772
Defining a Monopoly Airline Route	772
Defining a Reasonable Fare	773
Access Requirements	775
10. Conclusions	780
Endnotes	783
9 Constitutional Jurisdiction Over Transportation: Recent Developments and Proposals for Change	791
Patrick J. Monahan	
I. Introduction	791
II. The Existing Constitutional Framework	793
Jurisdiction Over Works and Undertakings	794
Interprovincial Works and Undertakings	797
Integration of Local and Interprovincial Undertakings	799
Works Declared for the General Advantage of Canada	801
Peace, Order and Good Government	802
Other Classes of Subjects	803
Regulatory Framework	806
Future Litigation	809

III. Proposals for Constitutional Change

The Practical Impact of Existing Constitutional Arrangements

Proposals for Constitutional Change

IV. Conclusion

Endnotes

10 Transportation Obligations and the Canadian Constitution

Patrick J. Monahan

I. Introduction: Transportation and the Canadian Constitution

II. The Confederation Era 1867–1873: Transportation

Undertakings as the Instrument of Political Union

A. *British North America Act, 1867*

B. Western Expansion: The *Manitoba Act, 1870* and the *British Columbia Terms of Union (1871)*

C. The *Prince Edward Island Terms of Union (1873)*

III. Making Good on the Obligations, 1873–1945:

Transportation Undertakings as an Instrument of National Policy

A. The Intercolonial Railway

B. Building the Canadian Pacific Railway

C. The Prince Edward Island Ferry

D. Summary: Political Over Legal Enforcement

IV. Completing Confederation 1949: Transportation

Undertakings in the Newfoundland Terms of Union

V. The Present Status of the Undertakings: The Movement

from Political to Legal Enforcement

A. The Prince Edward Island Ferry

B. The Prince Edward Island Railway

C. The Newfoundland Transportation Obligations

D. The Vancouver Island Railway

VI. Conclusions

Endnotes

PREFACE

Volumes 3 and 4 present a selection of research studies prepared for the Royal Commission by its Research Division staff and by various authors under contract. Volume 3 includes: historical overviews and general surveys related to transportation objectives; studies on subsidies, pricing and competition; and a discussion of institutional issues. Volume 4 includes: applied analyses related to determining the cost of transportation; industry studies of the air, bus and rail modes; and studies on travel demand, taxation and technology.

The historical overviews comprise two studies. The first, by D.R. Oworm, entitled "Icons and Albatrosses: Passenger Transportation as Policy and Symbol in Canada," examines the evolution of transportation in Canada, with particular emphasis on rail and roads. The second, by George W. Wilson, entitled "U.S. Intercity Passenger Transportation Policy, 1930-1991: An Interpretive Essay," provides a survey and a critique of U.S. transportation policy over the last 60 years.

Two of the general surveys, namely, that by Robin Boadway, entitled "The Role of Equity Considerations in the Provision and Pricing of Passenger Transportation Services," and that by David W. Slater, entitled "Transportation and Economic Development: A Survey of the Literature," discuss issues related to the inclusion of equity or economic development as objectives for a passenger transportation system.

The studies that discuss subsidies, pricing and competition issues include those by Trevor D. Heaven, entitled "Subsidies in Canadian Passenger Transportation"; David Gillen and Tae Hoon Oum, entitled "Transportation Infrastructure Policy: Pricing, Investment and Cost Recovery"; John Blakney, entitled "Competition Policy and Canadian Passenger Transportation"; and Keith Acheson and Don McFetridge, entitled "Controlling Market Power in Weakly Contestable Canadian Airline Markets." Federal-provincial institutional issues are discussed in two papers by Patrick J. Monahan, entitled "Constitutional Jurisdiction Over Transportation: Recent Developments and Proposals for Change" and "Transportation Obligations and the Canadian Constitution."

The applied analyses in Volume 4 include three studies on the cost of transportation. These are "Transportation Infrastructure Costs in Canada" by Ashish Lall; "Road Costs" by Fred P. Nix, Michel Boucher and Bruce Hutchinson; and "Environmental Damage from Transportation" by VHB Research & Consulting Inc. Of the industry studies, that by Steven A. Morrison, entitled "Deregulation and Competition in the Canadian Airline Industry," and that by Ron Hirshhorn, entitled "The Effects of U.S. Airline Deregulation: A Review of the Literature," relate to the air mode. The bus mode is addressed in "An Analysis of the Canadian Intercity Scheduled Bus Industry" by Richard Lake, L. Ross Jacobs and S. T. Byerley. The rail mode is addressed in the study by Charles Schwier and Richard Lake, entitled "VIA Rail Services: Economic Analysis," while airports are considered in "Airport Investment and Pricing Policies" by A. Cubukgil, S. Borins and M. Hoen.

Volume 4 concludes with studies on three further topics. Travel demand is addressed in two studies. The paper by Eric J. Miller and Kai-Sheng Fan, entitled "Travel Demand Behaviour: Survey of Intercity Mode-Split Models in Canada and Elsewhere," is a general survey of demand modelling. It is complemented by Richard Laferrière's study, entitled "Price Elasticities of Intercity Passenger Travel Demand," which calculates various elasticities of travel demand from several models on comparable bases. The impact of taxes on the cost competitiveness of Canadian intercity passenger transportation carriers, both intermodally and with U.S. carriers, is addressed in "Differential Taxation of Canadian and U.S. Passenger Transportation" by Ken McKenzie, Jack Mintz and Kim Scharf. Finally, a discussion of general technology issues and of prospective technology relevant to Canadian intercity passenger transportation modes over the next 25 years is provided in "Notes on Intercity Passenger Transportation Technology" by Richard Lake.

The contribution of those who participated in the editing and translation of all of the four volumes of this report was acknowledged at the beginning of Volume 1. In addition, the Royal Commission staff was ably assisted in the editing of Volumes 3 and 4 by PMF Editorial Services Inc.

ICONS AND ALBATROSSES: PASSENGER TRANSPORTATION AS POLICY AND SYMBOL IN CANADA

D.R. Owram*

February 1992

1. INTRODUCTION

This essay looks at the evolution of Canadian passenger transportation with a specific question in mind. Why is it that Canadians take such a proprietary, even affectionate, interest in issues of long-distance transportation, especially when railways are involved? Over the approximately one and a half centuries of Canadian history in which we have had a reasonably systematic system of passenger transportation, there have been massive changes in the technology and economics of transportation and accordingly in the way in which the public travels. Various periods have seen the dominance of differing forms of transportation. In the pre-railway era, water travel was by far the most efficient and comfortable. Beginning in mid-century, railways provided a fast, competitive means of travel by land and soon dominated both passenger transportation and land-based freight. In the 20th century, the arrival of cars made travel more feasible on roads, which until then had been a secondary means of travel. As this new technology was perfected and increasing amounts of money were spent on the road system, the car initially provided competition for and then largely displaced the train as a

* Department of History, University of Alberta.

means of passenger travel. Air travel, which arose most recently, further displaced rail travel, especially on those longer distance runs where cars had not yet captured the market.

Water, railways, roads and air all form important segments of Canadian passenger transportation. Of them all it is the railway that seems to possess the greatest symbolic significance. It rose in the mid-19th century as the new technology of the industrial revolution and soon towered over other forms of transport. Just over a century later it began to lose its central position, especially in the area of passenger transport. The number of passengers peaked in 1944, and thereafter an erratic but irreversible decline began. By the early 1970s railway travel was becoming increasingly unfamiliar to the average Canadian. Yet, despite all the changes, railways had an enduring, even mythic, hold upon Canadians.

People in many nations show a special affection for and interest in their public transportation systems, particularly railways. Canadians seem more fervent than most. From the time that railroads first appeared in North America until the present generation, books, films, songs and poetry have glorified the railway and its place in Canada's destiny. Indeed, when Gordon Lightfoot recorded "The Railway Trilogy" in 1962, he may have made Canada one of the few countries in the world to have a popular musical hit that glorified railway construction. That this was so is testimony to the way in which railways have become a part of popular culture in this country. A Grade 7 textbook of the 1950s reasonably asked the question, "Have you ever crossed Canada by railway?" Even if you had not, the imagery portrayed in the book would likely have been familiar. There was talk of the "kindly porter," "the snug upper berth," and the scenic ride through the "famous spiral tunnels." The passage concluded with the statement that "the road of steel has tied our provinces together."¹

To explore why such imagery has been so important to the national culture is a complex undertaking. This study attempts to understand our national concerns about passenger transportation by tying together three aspects of its development. First, we will look at the evolution of transportation systems and their importance in Canadian history. Second, we will pay special attention to government policy toward both transportation generally and passenger transportation specifically. Finally, we will discuss how people perceive the various modes of passenger transportation and their roles in Canadian society. Obviously, the three themes are intimately connected.

A great deal has been written on Canadian transportation. This is hardly surprising, for transportation is technology overcoming the tyranny of geography. This being the case, the very geography of Canada has determined the central role of transportation in the Canadian psyche. Geography, it has been argued, has always been central to the Canadian sense of identity.² Canada is the second largest nation in the world, and yet it has a small and scattered population. Thus, both the diversity of Canadian geography and the sheer magnitude of it have been inescapable throughout history. A vast nation with a small population means that transportation networks are going to be expensive relative to likely traffic. Yet the very distance involved in the geography of Canada makes efficient transportation all the more necessary.

This unfavourable distance/population ratio has resulted in three of the constants of Canadian transportation history. First, transportation—especially long-distance transportation—has been a more important issue in Canadian politics and economics than in many other countries. It is usually one of the items near the top of the public agenda.

Second, the distance/population ratio has led to a paradox that has confronted politicians, the public and businesses throughout history. In numerous instances a particular transportation project has been considered essential for the community or region involved. Yet, all too often, the small population and large distance meant that there was little or no likelihood of the finished work generating sufficient revenue to be economically worthwhile to any investor. The only way to bridge the gap was for government or governments to make the investment in the name not of profit, but of social welfare (real or perceived). As long ago as 1917 economist D. A. MacGibbon summed it up when he said, "Transportation policy has not been viewed merely from the standpoint of commercial and economic convenience. In the final outcome these national ideas have dictated the course of development."³

Third, as will be discussed later, transportation in Canada has come to be seen more than in most countries as central to the very existence of the nation. Although both Canada and the United States used the railway to tie together east and west, the links in Canada are more crucial and more vulnerable. Canada, with a smaller population, has always felt a strong pull toward the larger economy to the south. Given the sharp geographic division between the regions of Canada as well as the artificial nature of much of the Canada-United States border, an effective east-west transportation

system has from colonial times been thought crucial to Canada's continued independence. The alternative and geographically natural series of north-south interconnections seems to point to Canadian fragmentation and continental integration.

Geography, then, has created particular difficulties for the development of an efficient Canadian transportation system. This is well known, but equally important is how that difficulty has affected the public attitude. To put it briefly, subsidy has been a necessary part of Canadian transportation from the time mass transportation began. The Canadian public, therefore, is understandably conditioned to look upon transportation not as a series of profit-loss decisions by entrepreneurs but as a public good. That public good argument, moreover, has so far focussed more on trains than on air-planes or cars. This is likely because railways have been central to Canadian nation-building and because it is railway passenger travel that is currently threatened. This mood could, however, also apply to other means of transportation if they too seem both vital and threatened.

A great deal has been written about transportation in Canada, but relatively little on the history of passenger transportation. There has been no general study on passenger transport. Nor have those studying specific areas given this topic a great deal of attention. Instead, the emphasis has been on the construction of infrastructure or the movement of goods. Passengers do appear in studies, but there is little sustained description or analysis. There are exceptions to this rule. The emphasis on freight occasionally shifts during points of competition or crisis in the passenger area. Also, post-war developments of a large-scale air industry, which is more passenger dependent than railways, has tended to focus more attention on passengers. Still, that industry has had relatively few scholarly histories, and the passenger area, though important, remains thinly covered.

Part of this neglect results from the overwhelming attention paid to freight issues. Carriers were interested in the revenue passengers could generate, but the major decisions on construction, route, canalization or whatever were made with an eye to traffic in general, without any specific emphasis on passengers. For this reason it is difficult historically to distinguish between passenger and non-passenger issues. Very often they are part and parcel of the same concern. Even after works were completed the emphasis was often on freight. Only in the second half of the 20th century has passenger

traffic become a significant political issue. In contrast, controversies over freight rates go back much farther and, as the history of the Crow rates indicates, have been politicized for a very long time.⁴ Only by the 20th century were there consistently distinct corporate policies toward passenger and freight services.

2. THE ERA BEFORE RAILWAYS: TO 1850

Three characteristics mark the pre-industrial era in Canadian transportation. First, features of terrain and climate governed transport, for neither the technology nor the capital existed to significantly alter natural conditions. Second, because of this, transportation by water was much speedier and economical than that by land. By far the greatest movement of passengers and freight was along the natural water routes of North America. Third, toward the end of this first phase, the more advanced technologies of Europe began to have an effect on the transportation infrastructure of British North America. Steam power and the modification of the waterways by dredging and canals both improved the transport system and hinted at the more extensive applications of technology that were to become practical in the near future.

The Europeans who came to settle in North America were challenged by the vast and unknown landscape. Dense forests, ranges of hills and low mountains, seasonal freeze-ups and spring thaws all slowed travel. In the first centuries after their arrival Europeans did not have the population, the capital or, for the most part, the technology to significantly alter that landscape. As with the natives with whom they worked and competed, the Europeans had to operate within the North American environment. Fortunately, as any map of Eastern Canada indicates, the early colonists also found the landscape a useful ally: the strings of lakes and rivers provided an unparalleled access to the interior of the continent. The waterways thus became the highways of the colonial era. For more than two centuries the transport of both people and goods largely followed the rivers.

In contrast, travel on land was difficult and slow. Even at their best, roads in pre-Confederation Canada were inferior to waterways as a means of moving people or goods. Also, practically all the roads of this era were seasonal. Good winter and summer travel was possible, but spring-break-up

or the transitional fall frosts and thaws quickly rendered roads impassable. These facts shaped road policy. Because land transportation was so difficult in the forested eastern part of North America, settlement clustered around water and looked to water as a means of transportation. There was, for example, no road outside of Quebec City until the 1670s, more than half a century after settlement. Not until the 18th century was there really a system of roads or a structure of government to develop them.⁵ Even then they remained feeder systems for the most part, directing people and goods to the more efficient waterways.

Thus European settlements grew along the waterways, and European travel, as with native travel, followed those watercourses wherever possible. New France, whose settlers explored and developed trade far beyond the confines of the St. Lawrence Valley, has been characterized as a river empire.⁶ By the 18th century this empire reached west to the Prairies and south to the mouth of the Mississippi. This land was not transformed or conquered, and certainly no large-scale transport system was imposed upon it. Instead, the French operated through alliances with powerful tribes, and their travel in the region was on a small scale and conformed, as always, to the waterways. The Europeans had no technology suitable for such wilderness travel, and thus they adopted the natives' lightweight birchbark canoe. The great advantage of this was that it could be hauled over the numerous portages without great difficulty.⁷

By the 18th century European transportation technology had been introduced, at least in the more settled regions of the St. Lawrence. Flat-bottomed bateaux and local schooners operated between Quebec City and Montreal. Later, after the British conquest of 1763, the Durham boat (a keeled boat) joined the bateaux and schooners. These technologies extended westward to Lake Ontario by the late 18th century as the Loyalists and other immigrants began to settle the region that, in 1791, became Upper Canada.

Within this framework of small craft and a thinly populated continent, passenger transportation was a non-specialized and largely makeshift activity. A barge owner heading upriver or along Lake Ontario would sell space on his unsheltered barge for either freight or people as demand might have it. A crude tent might be erected, used by both owner and passengers, and food would be supplied as a part of the fare. None of this was fixed, however, and there were no formalized agencies or companies

engaged specifically in the movement of passengers until around the turn of the 19th century. Government, for its part, had few regulations and no investment as such in the provision of passenger transport. Moreover, even this rudimentary form of travel was limited in two ways. First, it was seasonal. When the river froze (as it did for four months of the year), those who wished to travel had to use sleighs (where they could get through the woods) or snowshoes. Needless to say, long-distance travel in winter was undertaken only under the most urgent circumstances. Second, even under ideal conditions, water transport was slow, especially against the current.

In the first years of the 19th century, passenger transportation began to change. For one thing, the growing population of Lower Canada and the rapid development of the frontier colony of Upper Canada increased demand for organized travel. Entrepreneurs saw that responding to this demand was potentially profitable. Moreover, the harnessing of steam, which was such an essential part of the industrial revolution then occurring in Great Britain, soon affected North American transportation.

In 1807 Robert Fulton demonstrated the practicality of steam power by successfully taking his ship, the *Clermont*, upriver from New York to Albany. Within two years a British North American launched the *Accommodation*. This initial craft was small and not really able to handle the currents of the St. Lawrence. A more powerful craft followed within a couple of years, however, and by the 1820s steamships were common on the St. Lawrence and Lake Ontario. By the 1830s they could also be seen on lakes Erie and Huron. Though steam engines were not yet perfected, these ships provided a degree of freedom from the current and the weather: they could work against the wind. Also, in terms of travel time, they brought points closer together. It was possible by the 1820s to travel from Montreal to Quebec City in about 24 hours. By 1835 a trip from Kingston to Toronto took about the same time, at least under good conditions.

Steamships not only improved the technology of passenger transport but also were of a scale that encouraged some specialization. These ships were not the small craft of just a few decades before, picking up a few passengers as demand warranted. The largest steamships could transport up to 900 people,⁸ and travel on this scale had to be scheduled to let customers know of sailings. By the 1830s formalized passenger transport existed

between Quebec City and Niagara, with (more or less) regular schedules, fixed fares, sometimes different classes of accommodation and (more or less) comfort.

As both trade and passenger transport became formalized and larger scale, government became increasingly involved, passing various regulations to ensure the safety of steamship passengers. Moreover, contracts to carry the Royal Mail could provide extremely important subsidies to steamship owners seeking to establish a profitable system along the St. Lawrence-Great Lakes system. The government also became involved in another way, one that established a precedent for future activities. Before long, those operating the ships and those travelling or trading along the St. Lawrence began to look to government for help in developing the infrastructure of the transportation system.

The St. Lawrence system was a superb natural highway into the interior of the continent, compared with any available elsewhere along the eastern seaboard. It was far from uninterrupted, however. Montreal and Lake Ontario were separated by a series of rapids. From the time that Jacques Cartier had been prevented from continuing his quest for China by the first of these (hence La Chine or Lachine rapids), travellers had had to make a series of time-consuming portages. Farther inland, at the escarpment separating lakes Ontario and Erie, was the spectacular break of the Niagara Rapids and Falls. Rapids on the alternative route into the interior, the Ottawa River, made travel difficult in that direction.

These rapids were an inconvenience, but for the light canoes and small-scale trade and transport common before the late 18th century they were tolerable. Portaging was simply part of the cost of travelling and trading in the interior. By the 19th century, however, the introduction of larger craft made the obstacles more costly. Moreover, the interior was no longer merely a sparsely populated hinterland. Upper Canada was growing quickly. From a practically unpopulated region in 1780 it grew within two generations to be the largest of the British colonies in North America. Already, by the end of the war of 1812, there were approximately 100,000 people there, and by 1830 that figure would more than double. Demand for cheap and efficient passenger and freight traffic was growing apace.

The result was a clamour for canal construction. The detailed history of this first great transportation enterprise has been told elsewhere, and only the

most basic details need to be repeated here.⁹ The effort began shortly after the war of 1812, with two separate ventures. One was in Lower Canada, where a group of businessmen decided to construct a barge canal around Lachine Rapids. The other was in Upper Canada, where local businessman William Hamilton Merritt planned the Welland Canal between lakes Ontario and Erie. Both were constructed, though neither their quality nor size was very satisfactory in the early years. Then, in the 1820s the British government provided assistance to the Canadas by constructing the Rideau Canal, linking the Ottawa River to Lake Ontario. By the 1830s it was thus possible, with the right type of vessel, to travel without portaging from Montreal to Lake Erie. The canals were often shallow, with as little as a five-foot (1.5 m) draft, and the route via the Rideau Canal was circuitous.¹⁰ In fact, for practically all freight and for most passengers, transshipment was still the norm. The small craft suitable for the canals were not suitable for the open waters of Lake Ontario.

In 1840 the two Canadas were united, and with improved fiscal capacity as well as a loan guarantee from Great Britain, they made a major effort to canalize the remaining rapids between Montreal and Lake Ontario and to deepen and improve the Welland and Lachine canals. By 1848 there was uninterrupted navigation to a depth of nine feet (2.7 m) from the ocean to Windsor. This, in combination with the refinement of steam-powered vessels that had taken place by the 19th century, revolutionized passenger travel. The passenger who could afford the fare was now able to travel in comfort by water to all of the major centres of Canada West (as Upper Canada had become in 1840). Moreover, the time of travel had been cut, though more important was the relative cheapness of moving heavy freight.

The canal era was a significant one in colonial Canadian history and was one of the key events marking the transition from a pioneer to a settled economy in central Canada. What is especially important for this study, however, is that the construction revealed two important and related paradoxes. First, the cost of canal construction demonstrated the already mentioned paradox brought on by Canadian geography. As a land of vast distances, Canada very much required an efficient transportation system. Yet the very distances involved and the small population made it extremely difficult for businessmen to undertake such works successfully. Though investors, businessmen and government officials might believe that transportation works were crucial, the necessary funds were still difficult to come by.

Local capital was scarce, and international capital was hard to attract for the very sensible reason that many of the works, however necessary, were unlikely to pay the cost of construction.¹¹

This was the case with the Welland and Lachine canals. The Welland Canal Company and the Lachine Canal Company, both chartered as private companies, quickly found that they could not raise sufficient funds to do the job. As early as 1819 the Lachine Canal, which had already been subsidized by both the colonial and the imperial government, was taken over by the government of Lower Canada. In Upper Canada the Welland Canal went through a more tortured process of acquisition, but the final result was the same. As early as 1824, wrote the historian of that canal, the "Welland Canal Company was already degenerating into a privately controlled institution for the disbursement of public funds."¹² The disbursements continued, but in a precedent that would echo through other financial crises the government also began to assume more and more control. By the later 1830s it was the largest shareholder, and by 1841 it had assumed direct control of the work.

By the time of the union of the Canadas in 1841, the government owned and operated the canal system. It was built, but not because it was seen as a paying proposition. Indeed, the construction and operation of the canals were always money-losing propositions for the government. Rather, the system was built because the benefits to the population were thought to make the works necessary. This raises the second paradox. The governments were only a little more able to afford the transportation systems than private enterprise. The canal systems imposed a severe fiscal strain on the Province of Canada. Before the union the Welland Canal had nearly bankrupted Upper Canada. The fiscal strain imposed by expensive transportation projects was not confined to the colonial era. As discussed below, it continued well into the 20th century, and echoes of it linger still.

There are many reasons why government allowed itself to become so deeply entangled financially. First, as various studies have shown, the dominant political groupings of the colonies were commercially oriented. They all saw growth as the key to the colonies' future and often, not incidentally, to their own.¹³ Second, given this assumption, government was trapped by the same paradox that affected society as a whole. If the work was essential, how could it be abandoned? Third, this problem had an extra dimension

at the government level. The canal system had always been promoted as an endeavour that was both progressive and patriotic. This was the way in which Canada could compete with the United States for the traffic of the mid-west. To fail to complete the system could jeopardize not only the national prosperity but also the national existence.¹⁴ It was an argument that was to appear many times in the future.

The burden of canals had nearly bankrupted the governments of colonial British North America. Then, just as the canals were finished and as the government undertook the ongoing cost of operating and improving them, a new and vastly more expensive transportation technology made its appearance.

3. LAND TRAVEL DURING THE GREAT RAILWAY ERA: 1849-1920

There was no parallel on land to the significant improvements wrought before 1850 along the waterways of British North America. Yet land travel was a necessity. Settlers needed access to their farms and had to carry their goods to market. For those on the fringes of civilization, access to decent roads was critical, from both an economic and a social viewpoint. Bad roads were ongoing matters of complaint in the pre-Confederation period. As early as 1799, grievances in Upper Canada were focussed on the state of roads. As one official document concluded, "The present wretched state of the means of communication between the several parts of the Province" almost paralyzed commercial and government activities much of the year. The government, it was urged, should take immediate steps "to remedy so serious an evil."¹⁵ The evil was not so readily overcome however. In 1819 a campaign of complaint against government focussed on the inadequacy of land communication.¹⁶ The 1837 rebellions in Upper Canada were in many ways a back-settlement rebellion against the dominating forces on the lake-front, including rebellion against the indifference of the colonial elites toward the road system.

By the 1840s the mood of rebellion had faded, but the farmers' sense of grievance remained. The elaborate canal plans of that decade brought a protest from those who warned that only a few miles inland farmers "lived in a country without practicable roads at a distance from society and civilization." There, because of the lack of effort and direction by government,

"they found themselves shut up in the gloomy prison of the eternal forest, excluded from the means of education for their families, from the consolation of public worship, from the blessings of civilized society. But they could not escape."¹⁷ Such imagery rang most true to those inland settlers who, in the 1840s, were pushing north toward Lake Simcoe or westward from London. Even for those in more settled areas, in Nova Scotia and Canada East as well as on the frontier of Canada West, the rhetoric was far from foreign. There were exceptions, but roads in British North America generally remained primitive even a decade or so before Confederation.

Aside from the obvious fact that roads were not very good, however, it is difficult to discern any clear stages in road policy or development before 1850. For one thing, the records of the pre-Confederation road system are sparse and imprecise. This is partly because the road is a much more amorphous concept than, say, is a railway. A railroad either exists or it does not. The same cannot be said for roads. In the years before the car, the trail, suitable only for people on foot and perhaps in existence long before European settlement, became a more defined route passable at least part of the year by horseback. Portions of the route then might be widened and improved to allow carts and wagons and ultimately be surfaced with one of the available technologies of the era. Bridges would replace fords. In all of these considerable changes, however, it is impossible to say for sure whether a new road was created. Yet this is exactly the progression followed by many of the land-transportation routes in the pre-Confederation era. The story of land travel from early New France through to the eve of the railway is one of gradual improvement rather than of any great technological breakthrough such as occurred with the railway or of any vast government effort as was the case with the canal system. Roads evolved as the poor cousins in the transportation system as needs demanded and resources allowed.

Both the limitations of government and the marginal position of roads dictated how road construction and finance were handled in the colonial era. Roads were treated as a local matter. At the extreme this placed the responsibility for the maintenance of a road on the family whose property it fronted. Generally, however, local governments were responsible for roads. Those local governments could and occasionally did use taxation to support the road system, but the general practice was to employ statutory labour. First in New France and then in colonial British North America, legislation

required citizens to work for so many days a year on the roads of their locality. Those who did not have the inclination to do so could commute their statutory labour by means of a payment to the local government.

Contemporary and historical sources unanimously agree that the statutory system presented serious problems. First, the level of maintenance and construction on roads depended on the enthusiasm, skill and willingness of the individual settler or, when organized into crews, the competence of the pathmaster and his amateur crew. Second, the effort made depended on the ability of local officials to bring out the population to undertake the work. It was also difficult to ensure that absentee landlords did anything at all along their property frontage. Thus, the quality of an individual road varied considerably from property to property, much less from county to county.¹⁸

There were always exceptions to the rule that roads were considered local matters. As far back as the 1720s the government of New France had designated certain key routes “chemins royaux” and had used colonial funds to construct and improve them. They had ditches and were 24 feet (7.3m) wide, compared with the 18-foot (5.5m) local, usually undrained, road. Likewise, in the British colonies, the colonial governments got involved in developing what were considered the most important roads. Governor Simcoe of Upper Canada used military funds for work on Yonge Street and the Dundas Road. In Nova Scotia the early 19th century government provided assistance for the main road system, which ran outward from Halifax to settlements like Antigonish, Yarmouth and Shelburne.¹⁹ There were also the “turnpike trusts.” These were roads built and maintained by private corporations in return for the tolls received. The main government roads and the toll roads were more likely to be maintained sufficiently to allow wagon as well as horseback traffic, to have bridges over at least the deeper streams, and to be drained in such a way that they could be kept open for a greater period of the year. They were still seasonal and relatively primitive even decades after settlement began in the region.

By the 1830s or 1840s the passenger setting out on a journey between two relatively major centres could hope to travel a road passable for horse and, optimistically, for wagon. The road itself would have been cleared of trees and brush to a width of 16 to 24 feet (4.9 to 7.3m), though shortage of funds or indifferent statutory labour meant the brush continually encroached upon the road and, along with fallen trees and branches, provided recurring

obstructions. With very few exceptions, the surface of the road was earth and, of course, highly susceptible to weather conditions. Heavy rain, spring breakup and similar events could quickly render even the best roads in British North America impassable.

Even when the ground was passable, the traveller faced problems. First, there were many low-lying, swampy areas that even a minor rain would turn into a quagmire. In these areas the technology most favoured by road builders was the so-called corduroy road. This consisted of logs (sometimes halved but usually not) laid across the road. Corduroy roads provided support for horse and wagon and prevented vehicles from becoming stuck. The rough surface also slowed traffic and was notoriously uncomfortable. Corduroy roads were necessary and common. They solved the problem of mud but only reinforced the image of land travel as slow and uncomfortable.

Superior technologies were also available. The plank road, which had sawn planks rather than logs, provided a smooth, fast ride by wagon. It was used in parts of the lumber-rich areas of New Brunswick and the Canadas. Even at its peak in the 1840s, it constituted a small percentage of road mileage, however, and required continual maintenance. Unless properly repaired, a plank road would within a few years become impassable because of the rotting and breaking of the surface.

Best of all the technologies was macadamization. The macadamized road consisted of a base of large rocks, progressively smaller rocks above and then gravel on top. In use in Great Britain by 1815, the technology crossed the Atlantic to British North America by the 1830s. It became increasingly common through the next couple of decades, especially with toll-road companies that sought to advertise their superior product. The reality was somewhat different from the rhetoric. True macadamizing was very rare in pre-Confederation British North America. With the exception of a few turnpike trusts near Montreal, the great majority of macadamized roads were really only gravelled. The firm base was never installed because it was too expensive. As a result, the so-called macadamized roads in Canada had neither the durability nor smoothness of the properly macadamized road.²⁰ Even primitive Canadian macadamization was a luxury, however, reserved for the major roads near urban centres or along certain turnpike trusts. The travellers of the 1850s, for example, might journey out of town on a gravel surface, but within a few miles they were likely to find only bare earth, with corduroy used in low-lying sections.

Rivers and streams also provided serious obstacles. Bridges were expensive to build, and colonial authorities avoided construction if at all possible. Travellers usually had to ford streams. Larger bodies of water also prevented easy travel. Often roads went miles off line to a suitable ford. Where bridges had been built, spring floods or severe storms often washed them away. Not until local and provincial public works became more organized around mid-century was substantial improvement made in this area.

The long-distance traveller on even the best roads could not expect an uninterrupted, dry, passable road surface over a long journey. If at all possible, travellers sought alternative ways to reach their destination. The lack of road traffic only discouraged further investment or effort. The roads simply could not compete. Typical was the York (Toronto) to Kingston Road shown in Figure 1. Pushed through in the years after 1815, the road was the major connection for the two largest settlements in the colony. Yet it was impassable for several months each year, required numerous fords, and was always threatened by the encroaching forest. Even at the time of the union it was in part a mere horse trail through the forest, poorly maintained and of marginal importance to all but local travellers.

Despite the obstacles, formalized systems of passenger transportation did develop. By the mid-18th century, stagecoach lines had been established in New France. There was also stagecoach service from Halifax to Windsor in Nova Scotia in the 18th century.²¹ By the early 19th century there were regular coach services in the two Canadas and Nova Scotia and New Brunswick. Companies were chartered by the relevant legislature and in return for guarantees of service were often given the contract to move the mail. By the 1830s there was a well-developed stagecoach system running between established settlements. The Montreal–Quebec City run, for example, took place two to three days a week and could be done — under ideal conditions—in two days. But then each day involved 16 hours of bone-jarring travel!²² Less frequent were feeder services running from small inland pioneer settlements to urban centres.

An early 19th century guidebook was pitiless in its denunciation of the stagecoach system, arguing that the roads are “poor” and “the surface rough, the bridges wretched, and the attendance at the inns defective. . . .”²³ For a stage to average five miles per hour (8 km/h) over any distance was to do well. It is hardly surprising that the stagecoach system had little enthusiastic

Figure 1
THE ROAD FROM YORK (1830)



Source: National Archives of Canada, Negative No. 12632 (reproduced in Edwin Guillet, *The Story of Canadian Roads* (Toronto: University of Toronto Press, 1966), p. 38).
Artist: James Cockburn (1778–1847).

support. It did have advantages at certain times of year, but it was seasonally restricted by the condition of the roads. Once away from the lake or river, therefore, Canada's transportation system was slow, uncomfortable and dependent upon the weather.

In all these ways, improvement over the last 200 years had been marginal. It is against this background that the tremendous revolution brought by the railway has to be viewed.

The first locomotive with sufficient power to justify commercial use was George Stephenson's British locomotive, the *Rocket*, which covered 70 miles at an average speed of 15 miles per hour (24 km/h) in 1829. Unremarkable though such speeds might be even a few years later, they were revolutionary at the time. This 70-mile trip was accomplished at the highest rate of sustained speed that man had ever experienced, and almost immediately railway lines were appearing in Great Britain and elsewhere. It was perhaps this sensation of speed and power that so quickly captured the public imagination. Images of power and speed were especially appropriate to the mood of the Victorian era. As many studies have shown, the possibilities raised by the industrial revolution, the effective enfranchisement of the middle class by the Reform Bill of 1832, and other measures convinced societies of the Western world that progress was the natural destiny of humanity. For the British subject, unsurprisingly, the centre of that world's progress was at home.²⁴ The proof of the progress was, for all its problems, the industrial revolution. Technology, combined with the entrepreneurial instinct, had brought material improvement, moral uplift and the spread of political liberty. And there was no more potent symbol of the power and might of industrialism than the railway. The steam locomotive became a symbol for the accomplishments of the age.

Canadians fully accepted the symbolism of the railway and, of course, feared what would happen if they were left behind. Thus, despite the horrendous costs involved and the small population of the colonies, various enthusiastic newspaper articles appeared from the 1830s onward extolling the wonders of the railway and asserting the need for immediate construction. It was in 1849, however, that the classic Victorian image of the railway was published. In that year well-known Canadian engineer Thomas Keefer published a tract with the pretentious yet appropriate title, *The Philosophy of Railroads*. The pamphlet was widely distributed, going through three editions in English and one more in French within four years.

For Keefer, the steel and steam technology of railways was an almost mystical means to free Canadians from the tyranny of nature. It is no accident that he begins his tract with a reference to the seasons: "Old Winter is once more upon us, and our inland seas are 'dreary and inhospitable wastes' to the merchant and to the traveller. . . ." The energetic activities of the colonizer and trader had been shut down. "The animation of business is suspended, the life blood of commerce is curdled and stagnant in the St. Lawrence." It did not have to be that way, for the Americans, always a powerful point of comparison, remained unaffected by winter. "Far away to the South is heard the daily scream of the steam-whistle — but from Canada there is no escape: blockaded and imprisoned by Ice and Apathy."²⁵

This contrast between progress and backwardness, between action and imprisonment, continues throughout Keefer's pamphlet. For though there are facts and figures, pragmatic arguments and suggestions, the tract is above all a morality tale. Indeed, Keefer's work is quintessentially Victorian in its linking of material and moral progress. In the centre is the hypothetical village of "Sleepy Hollow." With the arrival of the railway a spirit is engendered in the village that "is not confined to dress or equipage, but is rapidly extended to agriculture, roads, and instructive societies, and finally exerts its most powerful influence where it is most needed — in the improved character it gives to the exercise of the franchise." For where the railway went, enlightenment followed. "Poverty, indifference, the bigotry or jealousy of religious denominations, local dissensions or political demagoguism may stifle or neutralize the influence of the best intended efforts of an educational system; but that invisible power which has waged successful war with the material elements, will assuredly overcome the prejudices of mental weakness or the designs of mental tyrants"²⁶ The railway had gone from being an instrument of transport to a general agent for material improvement to, ultimately, the major civilizing force of the modern age.

Keefer's rhetoric may have been more inspired than most, but it both fit the age and reflected the dozens and then hundreds of pamphlets, monographs and newspaper articles that came out in the first phase of railway development in Canada.²⁷ Moreover, for all the hyperbole and propaganda, it is true that railways had a tremendous impact on the life of the average person. They transformed metropolitan-hinterland relationships, made inland transport cheap and rapid, freed travellers for the first time from the tyranny of weather, and quickly became the largest industrial concern in British North

America.²⁸ All of this reinforced the Victorian predilection to see railways as a metaphor for the progress of their society. Sudden bankruptcies, crooked schemes and the tremendous influence of railway barons on politics did not seriously affect this belief. For whatever the faults of the individual, the railway represented progress, and, it must also be remembered, the Victorian's notion of progress was a seamless web in which material and moral improvement were seen to be linked.

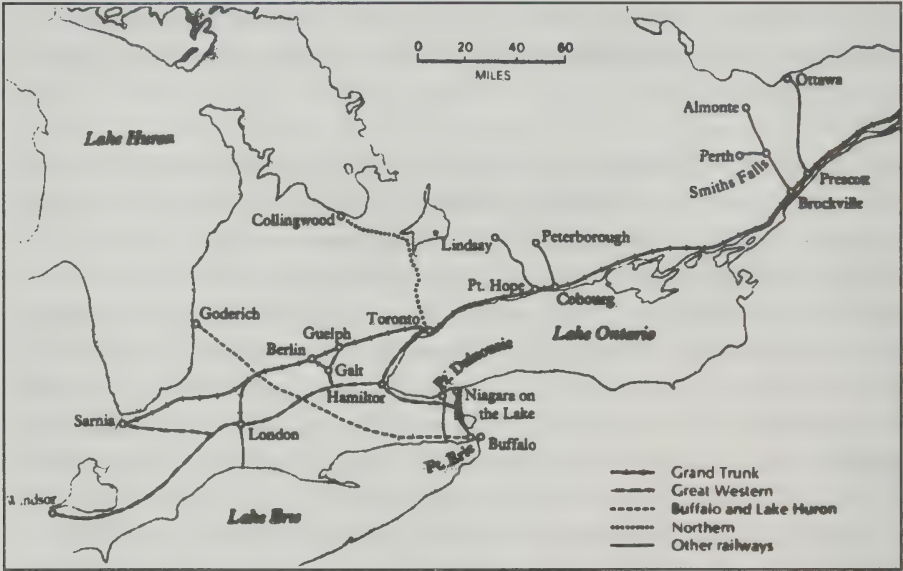
Keefer's pamphlet came at an appropriate time. The year it was published saw the beginning of the first serious railway construction in Canada. In 1849 there was only 66 miles (106 km) of railway in all of British North America. Eight years later there would be 1,800 miles (2,900 km). Millions of dollars (mostly British funds) would be invested (the Grand Trunk line alone cost \$67 million), and by the time it was all over many lines found themselves overextended and in debt. Receiverships followed close upon completion of construction.²⁹

As with canals, details of the first phase of railway construction and finance are well described in many works and need not be recounted here.³⁰ Basically, two types of railway were developed in this first period. One type was a feeder line, either to American lines or, more commonly, to water. Nova Scotia linked inland communities and the three coastlines by rail. In Canada West, railways like the Canadian Northern tapped the growing hinterland from Toronto to Aurora, Newmarket and ultimately Collingwood. The other type of railway was more ambitious. It was a trunk line designed not to feed the lakes but to provide a land-transportation axis from east to west. Both the Great Western and, most important of all, the Grand Trunk were this type. The latter, running from Sarnia to Montreal and then Quebec City, became the industrial giant of the age and a major force in the commerce and politics of Canada until World War I. By 1860 British North America had some 2,000 miles (3,000 km) of railways, most, as Figure 2 indicates, located in Canada West.

The relevant governments played a vital role in the evolution of these railways — in the negative decisions it made as well as the positive. Moreover, the decisions made during this first construction boom created precedents for government policy that lasted much longer than many of the railways then built. First, the Government of Canada determined that railways should *not* be directly built and operated by it. Likewise, the Government did not

assume responsibility for such support systems as railway stations like it did for harbours or, much later, airports. In all of this, the lessons of the canal era were crucial. Canals had led to tremendous financial strain, and their costs were nothing compared with the huge outlays needed for railways. At the same time, the Government was under the same pressure to develop railways as it had been earlier to develop canals. The railway, as Keefer's pamphlet argued, was crucial to the whole destiny of this young colony. Thus the Government sought to steer between the two extremes.

Figure 2
RAIL LINES IN CANADA WEST, 1860



Source: R. Cole Harris and John Warkentin, *Canada before Confederation: A Study in Historical Geography* (Toronto: Oxford University Press, 1974), p. 155.

Its strategy was to use indirect and occasionally direct subsidies. The single most important means was provided by the *Railway Loan Guarantee Act* of 1849. As the title implies, this Act allowed the railways to obtain guarantees on bond issues from the Government under certain conditions.³¹ Such a guarantee removed much of the risk to the investor and made it feasible for entrepreneurs to tap the large capital markets of Great Britain. The *Municipal Loan Act* of 1852 allowed cities and towns to draw upon provincial credit guarantees if they wished to support railway projects in their area.³²

In theory these acts provided a means of support for railway construction, at least under ideal conditions, without actually costing the taxpayers a penny. Not surprisingly, it wasn't that tidy. Railways borrowed heavily, and municipalities proved especially prone to railway fever. They had to, for the place of their community in the future hierarchy of trade and power depended on the railway. Many railways came crashing down, required reorganization or defaulted on their bonds. The Government had to step in on several occasions to rescue the railway or the overloaded municipality. By 1857 the first phase of railway construction was over, and for the next decade governments, companies and British bondholders sought to recover from the spree. All of this was reasonably predictable. Hundreds of millions of dollars had been spent in a region with a very small population and capital market. Still, the railways were in place, and this revolutionized Canadian transportation.

The railway affected the structure of industry in the colonies and the metropolitan-hinterland relationships, while the actual construction provided a major impetus to capital formation in the colonies. Those cities, like Toronto, that were the terminal points of major railways were the greatest benefactors. Extended hinterlands, new industries and the presence of considerable financial empires distinguished these cities from smaller pretenders and did much to establish the pattern of urban relationships that are still familiar today. For people as well as commerce the railways revolutionized life. For the first time it was possible to move as comfortably and rapidly on land as it had been on water. The farmer or merchant living inland in Galt or Barrie or elsewhere now had easy access to local towns or the larger cities. The hold of the seasons upon travel was considerably weakened and, with the improvement of snowploughs over the next decade or so, effectively broken. Year-round travel was now possible for the first time. In that sense Thomas Keefer's opening sentences about the hold of winter were entirely appropriate.

THE PROJECTS OF CONFEDERATION: THE RISE OF THE RAILWAY AS NATIONAL SYMBOL

Both railways and canals had been seen by governments in the colonial era as developmental necessities worthy of support. The projects had often had pro-British (that is, anti-American) overtones as well. It was at the time of Confederation, however, that transportation issues became linked closely to

nationalism. Confederation was accomplished in a mid-Victorian era, when faith in the industrial revolution and technological progress was high. One of the major impetuses behind Confederation was also physical expansion by the Province of Canada, which wished to "extend eastward to a port not in American hands and westward to new areas of trade and settlement."³³ The new Dominion was thus soon faced by the problem of continent-wide distances that only the railway seemed able to resolve.

The Province of Canada was not alone in seeking railway arrangements as a part of Confederation. In the Maritimes, early railway projects had brought the colonies near bankruptcy. The failure in the early 1860s of a projected railway link between Halifax and Quebec City had also left a feeling of bitterness, for the "Upper Canadians" were widely blamed for its collapse.³⁴ Those who supported expansion into the vast Prairie region realized that only the railway would make meaningful connection to the rest of Canada possible. Still farther west, British Columbia's entry into Confederation in 1871 made sense only if there was to be a transcontinental railway. Great Britain, having experienced the difficulties of moving troops overland during an Anglo-American war scare in 1861, was also supportive, especially of the Intercolonial Railway project.³⁵ Railways thus became linked with the very process of nation building. The Intercolonial Railway was actually written into the Constitution, the *British North America Act, 1867* (s. 145), making Canada perhaps the only nation in the world with a railway as part of its fundamental law. Once the precedent was established it was tempting to continue to invoke it. British Columbia's entry into Confederation was rewarded with a promise of a rail link to the Pacific; Prince Edward Island's, with a federal takeover of the bankrupt Prince Edward Island Railway.³⁶

The close connection between the political act of Confederation and the economic policy of railway development had important consequences. First, the imperative of national development led the Government to make exceptions to its policy of avoiding direct construction and ownership of the railway. The Intercolonial had to be built, but as one of Canada's senior engineers warned, "it offers no material for a flattering prospectus; we could not invite it to the attention of European capitalists as presenting an eligible investment for their surplus funds. But for the establishing of those intimate social and commercial relations indispensable to political unity . . . the railway is a necessity."³⁷ There was thus no choice for the Government but to build it itself.

Dependence on the Government brought problems. For example, the choice of route — the so-called northern route — saw purely commercial considerations downplayed in favour of military and political ones. Also, there were fears of patronage (for good reason, given the times), and a cumbersome, three-layered authority was set up: the chief engineer of the Department of Public Works, the Government and appointed commissioners. Finally, once the taxpayer's money was involved, decisions on quality of construction became political topics and led to a great deal of wrangling back and forth.³⁸ Not all the problems were resolved satisfactorily, but by 1876 the first through train ran from Halifax to Quebec City.

The second implication of this national project was that having been built it could not easily be abandoned when it proved costly to operate. The Government had hoped to get rid of it by sale or lease to a private operator such as the Grand Trunk. The Intercolonial was no more attractive to operate than to build, however, and it became a Government railway line, operating at a loss from the day it was opened. Before long the Government also acquired another railway, along with a province. Prince Edward Island had resisted Confederation but had not been able to resist railway fever. By the early 1870s the Prince Edward Island Railway had brought the small colony to the edge of bankruptcy, and Canada moved quickly to offer to bail them out: it would assume the inefficient and badly built railway in return for that colony's entry into the Dominion. By 1873 the deal was completed, and the Department of Public Works found itself in charge of yet another railway. It too operated at a loss.

The Government's involvements in these Maritimes railways were patently political bargains. Many people in the Maritimes since then, therefore, have seen adequate service for passengers and reasonable freight rates as not just another regulatory decision but a part of the basic constitutional package of Confederation. In the 1920s the powerful Maritimes Rights Movement sprang up to protest apparent federal indifference to a region undergoing economic decline. Much of the rhetoric and anger of this movement, as we will see, were focussed on Government management of the Intercolonial.³⁹ Likewise, during the more recent controversies over VIA Rail service, Maritimers pointedly argued that the Intercolonial was a part of the Confederation pact.⁴⁰

The scope and cost of the Maritime railways were nothing compared with the greatest railway project to come out of Confederation: the transcontinental Canadian Pacific Railway (CPR). The story of the construction of the CPR has been told on numerous occasions. Suffice it to say here that after some failed organizational attempts in the early 1870s, there was desultory and half-hearted construction, accompanied by energetic surveying in western Canada in the late 1870s. Then in 1881 the Government enticed the famous "Syndicate," headed by James Hill, George Stephen and Donald Smith, to undertake the work in return for suitable incentives. Remarkably, that work was completed by November 1885. Unlike the eastern railways discussed above, the Canadian Pacific soon became profitable and astounded its creditors by quickly repaying outstanding loans. It rapidly became the most important corporation in Canada and one of the major railways of the world.

Government policy toward the CPR was based on traditional forms; that is, the Government avoided ownership and gave subsidies instead. Some direct Government construction was undertaken in the 1870s by the Liberal administration of Alexander Mackenzie, but the major portion of the line was constructed on principles derived from earlier railway ventures in Canada and from the United States. Specifically, the Government offered the investors three types of subsidy. The first, borrowed from the Americans, was a land grant — the famous 25 million acres. The second was cash, initially \$25 million. The third was a series of privileges accorded the CPR: the remission of duty for construction materials, the promise of monopoly between the main line and the border for 20 years, and other lesser provisions. As construction went on and the company ran into financial difficulty, additional support was given.

There has been much debate over the years, and especially recently, about whether all of these subsidies were needed.⁴¹ It is unlikely the question can ever be answered conclusively, but from the point of view of this study that is less relevant than the simple fact of subsidy. For the subsidies created an ambivalence in the relationship of the CPR with the Government that persisted thereafter. On the one hand, the CPR is not a purely private work. The CPR is often seen by the Government and the public as having a particular public service role, resulting from the support it was given at the beginning. On the other hand, as the CPR has pointed out at various times, it is a private company. Once it completed construction, it repaid Government loans and

operated on a self-sufficient basis thereafter. As a CPR publication of the 1940s put it, these government supports should be seen as “payments for service rendered” rather than as subsidies at all.⁴² From this perspective the CPR has a primary duty to its shareholders to uphold its profitability and therefore to operate freight or passenger services as commercial needs dictate and not as an agent of social or political policy.

The most important aspect of the CPR, however, is the standing it has assumed in the Canadian psyche. No other railway, and perhaps no other commercial project in Canadian history, has assumed such an important place in the public consciousness. The CPR has become a legend and a symbol in both scholarly and popular history, and the nature of that legend says much about the values Canadians attach to their railway system.

The CPR was controversial from the very beginning. The early attempts to create a viable private company collapsed amidst the Pacific scandal of 1873. Along the way it brought down the government of John A. Macdonald. Then, when the Syndicate was formed in 1881, there was a great deal of controversy over both the principle of vesting such massive power in one group and some aspects of the subsidies, especially the notorious monopoly clause.⁴³ From potential investors came the warning that the railway could look forward to a “long and dreary season of unprofitableness.”⁴⁴ The controversy only intensified when the railway was completed. Almost immediately Manitoba attempted to subvert the monopoly clause of the charter, and farmers’ groups were protesting the high-handed attitudes of the railway before the 1880s were out. Resentment against the power and policy of the CPR flared regularly across the West. Freight rates, grain-elevator policy and the need for branch lines made the CPR a much-loved object of hate through Western Canada. Indeed, many of the agrarian organizations and reforms that came about over the years owed much to a common hatred of the CPR.

Even more powerful than this distrust of the corporation was a national enthusiasm for the railway itself. Titles like *The Romance of the Canadian Pacific Railway*, *Steel of Empire*, *Canada’s Great Highway* and, best known of them all, *The National Dream* reflect something of the romance and glory attached to the railway over the years. Even in the West the predominant writing on the railway has tended to distinguish between the corporate rulers, eastern and callous, and the railway itself — a vital lifeline on which westerners worked, travelled and depended for trade.

Part of the image of the railway has much to do with the sheer ambition of the project. As the advertising copy for the 1949 movie "Canadian Pacific" dramatized it, this was "The story of a railway that couldn't be built . . . but was."⁴⁵ This romantic image of the CPR began with contemporary accounts of the adventure and hardships involved in its construction. The treks through uncharted wilderness, the bogs of northern Ontario, the hardship and even the deaths all became a part of the glamour of the project itself. Equally, early travellers over the line — often before the line itself was complete — related details of the hard work, the danger and the adventure that surrounded the project.⁴⁶ The CPR, still desperate for money and respectability, did everything it could to encourage such writing. The travel of the prestigious British Association for the Advancement of Science over the line in 1885, for example, led to good publicity.⁴⁷ So too did the transportation of troops west to suppress the Riel rebellion in the spring of that year.

As time went on, the romance of construction was reinforced by a sense of nostalgia and the important fact that the dubious venture of the railway had in fact succeeded. As the Canadian Pacific survived and then prospered, repaid loans and made profits, the earlier uncertainty disappeared. The very decision to act became as romantic as the action itself. "It was in itself an act of splendid audacity for a people of less than four millions in number to start on the task of throwing a railway across an immense and almost uninhabited continent to the shores of the Western sea," wrote one enthusiast in the 1920s.⁴⁸ The actual construction, for its part, became a part of the great Canadian epic of taming the wilderness. When engineer Sandford Fleming witnessed the laying of the last spike he linked it to the earlier generations of fur traders, "the Mackenzies, Frasers, Finlaysons, Thompsons, M'Leods, MacGillivrays, Stuarts, MacTavishes, and M'Loughlins, who in a past generation had penetrated the surrounding mountains."⁴⁹ Time has not eroded the tradition, and many others since have emphasized the last spike as part of the exploits of men who conquered the wilderness. The official government plaque commemorating the last spike concludes, "A nebulous dream was a reality: an iron ribbon crossed Canada from sea to sea. . . . Here on November 7, 1885, a plain spike welded East to West."⁵⁰

There is yet another strand to the legend of the CPR. Victorian popular beliefs held that the individual could make a real difference through the application of hard work and will. The myth of the CPR was a celebration of both the Protestant work ethic and the industrial revolution.⁵¹ People like

Stephen, Hill and Van Horne were naturals for such a role. Daring to venture where others feared to go, they made themselves the most powerful businessmen in the land. They were the “captains of industry” who had carried through such an audacious act, who had triumphed over the doubters and the sceptics. They were the “men of vision” to undertake the task, to do it well and to complete it.⁵² In their lives and characters they became archetypes, symbols of the values and morals of an age. William Van Horne was not just an exemplary railway builder but a paragon of hard work and morality. During the Chicago fire, it was recorded, he had risked life and limb to remove material from his company’s offices. Then, black with soot, he returned home only to continue his heroic endeavours: “Reassuring himself of the well-being of his wife and her infant, he set to work very quietly and industriously to strip his home of everything, and more than everything, that could be spared.” Bedding and clothing were then sent to “the shivering refugees from the South Side who were camped in the park.”⁵³ In the more jaded world of the late 20th century the men of the Syndicate have stood up much better than most Victorian heroes. Their courage and competence seem to distinguish them from other old idols who were shoved aside as new sensibilities emerged.⁵⁴ Even in their less desirable attitudes the men of the Syndicate were romantic figures, “larger-than-life characters, who ruled with the kind of iron-fisted autocracy that may never again be possible.”⁵⁵

The adventure, the romance, the embodiment of hard work and even vision have all been assigned to the Canadian Pacific or its builders over the years. All of these characteristics are, however, merely embellishments upon the main symbolism of the railway — nationalism. As a CPR publication immodestly concluded in 1946, “All down the years Canada and the Canadian Pacific have marched ahead together. Indeed, to the world at large, the two have become synonymous.”⁵⁶ Self-serving though such a company statement might be, it is also a reasonable claim. For in both the scholarly and the popular portrayal of the CPR, nationalism has loomed large. This nationalist image, moreover, has survived to the present and may explain much of the uneasiness about the decline of the most visible of all railway activities, passenger service.

The national connection is, of course, based upon historical fact. At the time, both the Government and the CPR recognized that annexation and settlement of the North West was impossible without a railway. Equally important, however, was the evolution of a particular geopolitical analysis

of the nature of Canada. This viewpoint reinforced the premises behind the Canadian Pacific and became central to later popular as well as scholarly interpretations. This geographical aspect was summed up in its starkest form by the late 19th century intellectual Goldwin Smith. Canada, he said, was really “four separate projections of the cultivable and habitable part of the Continent into arctic waste.”⁵⁷ For Smith such geographical realities meant that Canada’s inevitable destiny was annexation to the United States. For most other observers the result wasn’t that inevitable, but the difficulties posed by the geography of North America were central to the analysis of Canada as a national entity.

Few assessments of the Canadian economy and nation ignored Canada’s distinct geographical regionalism. It was, however, the “staples school” of Canadian economists that turned physical reality into an elaborate and influential theory. Aspects of the theory were present in the works of pioneer political economist O. D. Skelton.⁵⁸ In the 1920s, W. A. Mackintosh of Queen’s University and Harold Innis of the University of Toronto made the theme explicit. The massive research work of the Rowell-Sirois Commission (1937–41) drew upon the work of a generation fully imbued with these ideas and incorporated it into its own assessment of the Canadian nation.⁵⁹

The staples thesis in its basic form posited that colonies grew to economic maturity through the exploitation of a series of staple products — furs, timber, wheat, minerals. Such staples, under certain conditions, provided the means for population growth, development of entrepreneurial skill, and capital formation. These things, in turn, permitted the development of a more complex economy. The formation and evolution of a staples economy were also seen to have a major role in national development. Commercial links developed in this manner determined imperial and then national boundaries. This was what explained Canada’s separate existence in North America. The staples exploited through the centuries had been linked into an East-West trade system with Europe (and especially London) as the metropolis. This East-West system overrode the disuniting forces posed by geography and tied the regions together. In a famous phrase, Mackintosh said that Canada existed “in defiance of geography.”⁶⁰ Innis, looking to the rivers as a precedent, disagreed, stating that it existed because of geography.⁶¹ All economists of the staples school, however, felt that railways were the technological instrument that allowed the East-West trade flow to continue into the modern age.⁶²

The distinctly unromantic prose of Innis and Mackintosh contained within it a dramatic historical theme. For in their interpretation the credit of creating and maintaining Canada went to those who in some way grasped the essential reality dictated by geography and shaped by national pride. Exploitation of staples wasn't just a matter of luck but of policy. Those who failed to understand or who deliberately stood in the way of the opportunities were seen as negative, narrow individuals, opposing destiny.⁶³ This was clearly sensed by those who, drawing upon the staples thesis, developed what became known as the "Laurentian Theme" of history.⁶⁴ Individuals with more flair for drama (and more readers) picked up the themes of Innis and Mackintosh and merged them with the long-standing romance associated with the CPR. Perhaps the best known of these individuals was Donald Creighton, the foremost historian of the years surrounding World War II. Creighton first picked up the vision of the east-west transcontinental vision from Innis, whom he much admired, and translated it into his own words in his 1937 work, *The Commercial Empire of the St. Lawrence*.⁶⁵ This work dealt with the years of furs and canals and thus predated the CPR. This was only the first part of the story as far as Creighton was concerned, however, and he brought the vision to fullest fruition in his masterful two-volume biography of John A. Macdonald. Macdonald and those who comprehended the vision of a transcontinental Canada became heroes. Confederation and the Canadian Pacific were made inseparable, and people like Stephen and Van Horne were idolized.⁶⁶ Creighton's work became classic, one of the best selling works of academic history ever written in Canada. As such it did much to reinforce and expand the association of the CPR with the national vision, especially as the work was published during a period of post-war nationalism and confidence.

Yet Creighton's readership pales when compared with that achieved by Pierre Berton's rendition of the CPR legend. Published in two volumes in 1970 and 1971, Berton's history of *The Great Railway* was a phenomenal success. In a country where a sale of 5,000 books can create a best-seller, his railway books sold more than 100,000 copies and are still in print. Furthermore, the works were later developed into a mini-series by the Canadian Broadcasting Corporation. Taking a romantic nationalist interpretation of Canadian history, Berton used a popular variant of the Laurentian school. Even the title of the first volume, *The National Dream*, summarized the symbolism of the railway:

Macdonald's vision of Canada did not stop at the Great Lakes; his dream was of a transcontinental British nation in North America — a workable alternative to the United States. To achieve this dream, the railway was a necessity, or so the Prime Minister insisted: it would stitch the scattered provinces and empty territories of the West together, as the government-owned Intercolonial was intended to do in the east; . . . it would forestall American expansion; it would be the spine of empire, an Imperial highway linking the British Isles with the Orient and avoiding the appalling voyage around the Horn.⁶⁷

Berton's books were so successful not just because of the popular name behind them or the readable prose. He has written many other books, before and since, and none have matched these in impact. What made them so remarkable was that he adopted a set of beliefs about Canadian development and about the importance of the railway to that development that was already firmly established in the Canadian mind. The fact that he wrote at a time when the national sentiment was still high in the wake of Expo 67 and at a time when railway passenger services seemed threatened simply accentuated the appeal.

Berton and Creighton have been singled out only because they are among the best known and best selling historians of the era. Numerous others could have been cited, and at some point a graduate thesis will dissect the components of Laurentian nationalism. What is important, however, is that a part of the standard wisdom of Canadian nationalism is that there is a close connection between the success of the railway and the triumph of the young Dominion. It is a lesson taught by the scholarly interpretations implicit in the national policy and staples thesis; it is taught in the textbooks of school children; and it is found in the many popular readings on railways. There is even the already-mentioned hit song by Gordon Lightfoot and an epic poem, "Towards the Last Spike," by E.J. Pratt. " 'No Road, No Union,' and the converse true / East-West against the north-south run of trade."⁶⁸

The CPR has from the beginning recognized the marketing and political value of such an image and uses it whenever possible. It has even hired popular historians to recount the daring, the vision and of course the nationalism of the railway. For example, John Murray Gibbon's commissioned work, *Steel of Empire*, published in 1935, links the railway not only to Canadian nationalism but also to the age-old European dream of a connection to

China. The first chapters deal not with the CPR at all but with the long-standing search for a route to the East, the exploits of the fur traders, and other incidents in Canadian expansion across the continent. The message is clear. The CPR was not just the whim of some Confederation-era politicians but the culmination of a centuries-old dream.⁶⁹

The nationalist imperatives that surrounded the construction of the Intercolonial and the CPR added a new dimension to the railway in the Canadian mind. From social or economic necessity it had now become a national necessity, even an icon. The additional elements of success, daring and romance only added to the appeal. For some Canadians the transcontinental links were a part of the Confederation package. Maritimers have long argued, for example, that the Intercolonial was "part and parcel of the contract of Confederation."⁷⁰ Even for Canadians who did not interpret provision of rail service as a quasi-constitutional arrangement, the railways and especially the Canadian Pacific were intimately connected with the founding of the nation. That connection was and is reinforced by the tremendous success of the project. Here in a dry, unemotional Confederation process was a heroic achievement on a world scale. The nationalist association, as we will see, spills over to all railways in Canada to a degree. The public's sense of attachment to railway travel, or to the possibility of railway travel, is very likely linked to their identification of successful, operating railways as a part of the national framework.

The railway now totally dominated land travel over any distance. In contrast, the development of roads languished. Indeed, as many later saw it, the emphasis on railways precluded any serious attention to roads. As a 1930s Ontario royal commission put it, "For many years they [roads] had been neglected in favour of the railways. . . ."⁷¹ This echoed a theme that had been around for some time. In the late 19th century a group of citizens came together to complain about the state of roads: "Had railroading and navigation by steam been deferred for another half century, we would have been advanced farther in the science of road-making."⁷²

Certainly the railway made the road even more marginal than it had been. For the railway was able to go inland, to penetrate beyond the headwaters and to go wherever sufficient commerce and passenger traffic existed. Railways became both the feeder systems to water and, as in the case of the Grand Trunk, competitors with waterways. Passengers travelling between

cities now had a fast, year-round and comfortable means of travel. As the rail system developed throughout Eastern Canada from the 1850s to 1880s and Western Canada from the 1870s through early 1900s people had fewer reasons to travel any distance by road. Even in the neighbourhood of cities the road came under challenge. Commuter trains and their short-haul analogues like streetcars meant that rail travel became the basis for early commuting and suburbanization.

It is difficult to make too many generalizations during this era of neglect. Still, a few general points are apparent. First, the Dominion government had a very limited interest in the Canadian road system. Whereas other public works, namely the canals, had been declared to be for the national benefit under section 92 of the *British North America Act, 1867* and though railways became objects of national policy and large subsidies, roads were ignored. Second, among the provinces and territories, local administration was the rule. Roads were left to counties or their equivalents to maintain and improve. Statutory labour, local tax levies and rare provincial subsidies were the basis of a rudimentary, underfunded system, one that saw maintenance on all but the most major roads conducted in just a few short weeks every year.⁷³

There was one exception to this general rule of indifference, though the exception is a spotty one. The federal government and some of the provinces were active on the frontiers of settlement. The development of colonization roads was a practice that predated Confederation and was continued for sometime thereafter in both Ontario and Quebec. Where crises of access developed, as during the British Columbia gold rush of the 1850s and 1860s or the Yukon gold rush of the late nineties, crash programs were often undertaken. When the effort was made the results could be impressive. Immediate imperatives in the 1860s saw the completion of the Dawson route, a mixed water-land route from the Lakehead to Manitoba, and the 385-mile (620 km) Yale Road, built to handle gold rush traffic in British Columbia.⁷⁴ Even in these exceptional instances, however, roads were viewed as poor substitutes for railways. Indeed, the assumption was that the road was a temporary measure, preparing the way for future railway development. The Dawson route was never thought to be a long-term solution and was abandoned after the CPR replaced it. It would be many decades before the Lakehead and Winnipeg were once again connected by road. Likewise, parts of the Yale Road were abandoned when the railway

reached British Columbia. In the Yukon the boom times of the turn of the century brought immediate plans for the replacement of the rough trails inland with a railway.⁷⁵

Finally, in contrast to the railways, roads played no part in the national imagination of 19th century Canada. They were, if a topic at all, one of complaint on a local level rather than of dreams at the national. This is reflected in the historiography. O. D. Skelton's book, *The General Economic History of Canada 1867–1912*, ignores roads.⁷⁶ As part of the massive 23-volume history of Canada published in 1914 by Adam Shortt and Arthur Doughty a special 112-page article on land transportation was commissioned. Roads got just over six pages of coverage before the article shifted to the really important story — railways.⁷⁷ Even as late as 1938 the standard study of transportation in Canada by G. P. de T. Glazebrook has nothing at all on roads between 1867 and 1900.⁷⁸ Every chapter in this book from Confederation to World War I has "Railway" in the title. Not until the 1960s, by which time the importance of roads had changed considerably, did anyone write anything significant about their history. As it is in history, so it was in the contemporary material: the monographs, government studies, public debates and other materials were mainly on railways or canals, not roads. Roads were referred to, of course, but only as incidents in travel memoirs — usually horrible, occasionally pleasant — and not as a significant element in the national vision. Roads remained at the margin of Canadian transportation policy between the arrival of the railway and the turn of the century.

PASSENGER TRAFFIC AT THE TURN OF THE CENTURY

It is extremely difficult, except in the most rudimentary way, to distinguish between policy aimed at railway passengers and that aimed at railway freight before the late 19th century. In the minds of governments and companies, development was designed to encourage the transport of both. Yet it is likely that the romance attached to railways over the years applies more to passenger travel than to freight. An individual's sense of connection with the railway has much to do with memories of trips on the railway to visit relatives or friends or to take holidays. It may also have much to do with family history, with parents or grandparents whose life in Canada began with a railway trip. For many more the railway was until recently the primary tie to the outside world, connecting isolated rural communities to the larger centres. It is therefore crucial to explore the evolution of passenger policy

and the passenger experience more closely. With the approach of the 20th century, railway and government policies changed, making it easier to trace this evolution.

Government involvement in railways was initially experimental and fluid, as might be expected: it was coming to grips with a new technology. Much of the pre-Confederation legislation was designed to control the financial and construction excesses of this powerful new form of corporation. Railway acts were concerned with destroyed animals, seized property, ruined fences. Legislation created a Board of Railway Commissioners to try to ensure railway compliance with the rules.⁷⁹ The railways, for their part, more or less did what they felt was necessary and then sought to present their case to government or the courts. In the actual operation of the railway, however, initial government activity was limited to two general areas. First, there were rules on passenger safety, an understandable intervention given the horrendous accidents that began to occur as soon as railways began operating in the 1850s. Second, there were rules to prevent discrimination, drawn from British precedent of the railways as “common carriers.” Under this principle all people are offered similar services at similar prices.

With Confederation, colonial precedents were carried forward. In particular, the concept of the common carrier was now well established, and the Railway Committee of the Privy Council was established under the 1868 *Railway Act* to oversee railway activities. This body of appointed politicians had the responsibility for everything from rules of signage to right-of-way privileges. Also within its ambit were rules for the safety and comfort of passengers. The committee was far from ideal as a means of railway regulation. It had neither the staff nor the expertise to undertake detailed investigations. Nor did it always comprehend the implications of the decisions it made. This was also an age of laissez-faire, and the tradition of regulating the common carrier contended with a desire to leave private business alone. Still, by the time the committee went out of business in 1903, a considerable body of statutes and regulations on passenger transport had evolved in Canada. The most important elements of these were incorporated into the new *Canadian Railway Act* (3 Edw. 7, c. 58). The Act also gave the newly appointed Board of Railway Commissioners power over such passenger issues as standards of the railway station, the conductor’s right to expel passengers and various safety measures. Further, in a precedent that could, if extended, apply to line continuation, the Act gave the Commissioners the

right to set tolls and to force trains to alter schedules to assist passengers in making connections between different corporations (s. 284). Thus, as early as 1903, the doctrine of the common carrier contained within it the principle that railway operations, and presumably profitability, could be intruded upon in the interest of not only public safety but also public convenience.

Parallel to this was a growing effort by the railways to attract passenger traffic. From the beginning the CPR understood both the dollar value and, perhaps more importantly, the positive publicity that went with satisfied tourists. The West was still the frontier, and the promise of good hunting and dramatic sightseeing and the opportunity to meet real Indians and cowboys attracted tourists from more eastern lands. In the years immediately after the completion of the railway it became common practice for affluent British tourists to make the trek across Canada and write about their experiences. The authors naturally sought, with greater or lesser skills, to convey a sense of romance and adventure, which only added to the aura of the Canadian railway adventure.⁸⁰ In the East the Canadian Pacific and the Grand Trunk issued guidebooks that talked of the glories of Niagara Falls or of fishing in New Brunswick.⁸¹ More will be said about the tourist trade below, for, although it was a growing part of the railway business, it was not until later that it became really important.

Success in passenger traffic as far as the CPR was concerned depended upon immigrants to the West and the spin-off traffic that resulted from their homesteading efforts. Between 1881 and 1914 the Prairie population increased from just over a quarter of a million to more than 1.3 million. It was the Canadian Pacific that brought most of these people westward. In the first full year of operation the line carried more than one and a half million passengers. By the early 1890s the figure had doubled to more than three million. When the wheat boom hit the Prairies at the turn of the century the numbers continued to rise, to more than four, then five, then seven million. By 1913, the last year of the boom, the CPR carried more than 12 million paying passengers.⁸² To handle this traffic the railway built special immigrant cars, advertised in Europe and arranged through journeys at cheap rates. And because they owned vast tracts of Prairie land as a result of earlier grants, the railways could run a service that handled the immigrants' needs from the time they embarked in Europe until they were settled on the land. It was a nicely integrated relationship. That original passenger trip could lead to land-sale revenue, and that sale could, within a couple of years, lead to

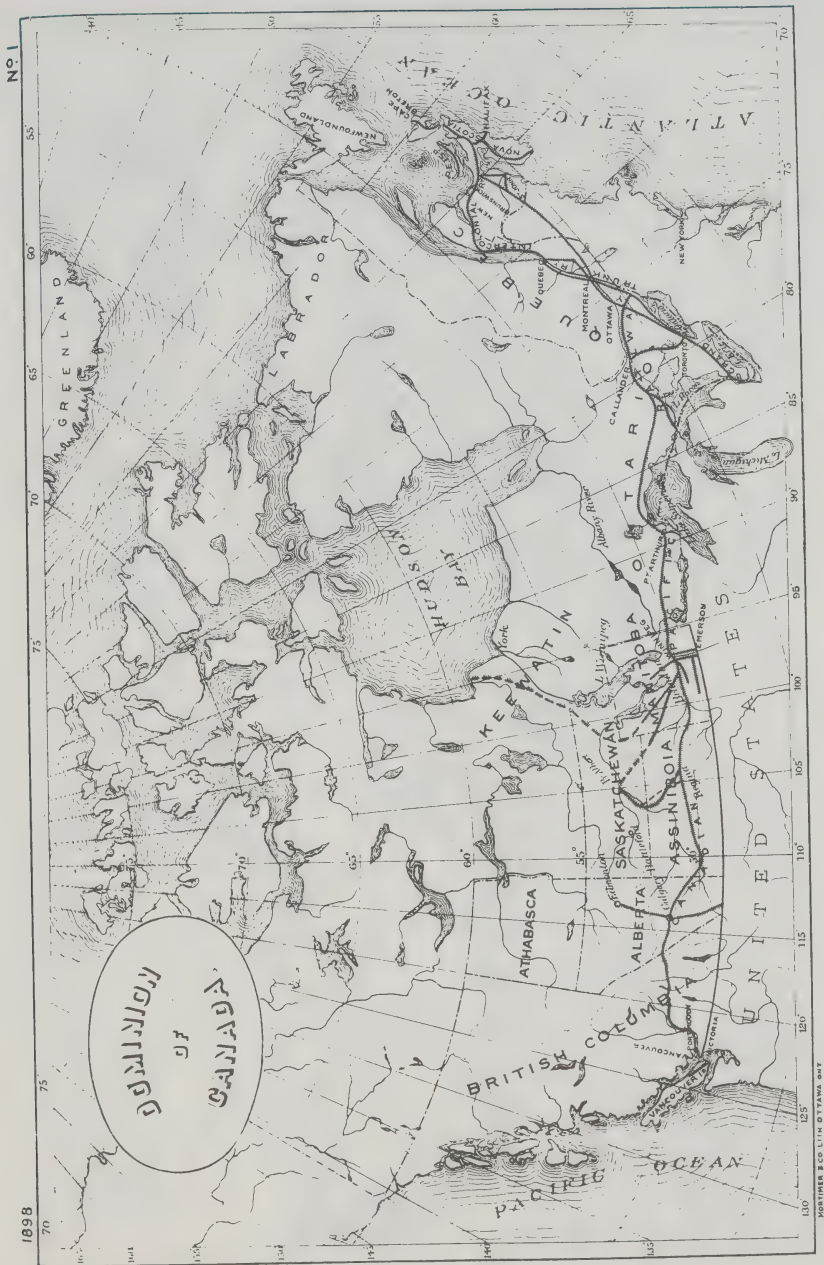
increased freight traffic as crops were exported and goods imported. Throughout the West the increasing population brought an ever larger number of commercial passengers to Prairie towns and cities. Thus the original cheap immigrant fare, profitable in itself, was repaid in many ways over the years that followed.

Obviously not all passenger traffic depended upon immigrants. In the East, where both the CPR and the Grand Trunk were major systems, the traffic was different. Most of the people using the trains were making round trips. Their trips were shorter and often involved major urban centres like Toronto, Montreal or Halifax. People might be travelling for business, for holidays or for visits with relatives. In many cases the passengers were commuters, travelling short distances from satellite communities to business in the city. Even in the West the nature of passenger traffic was evolving. In the first decades the immigrant traffic was central to Canadian Pacific's passenger totals. By the early 20th century the immigrant traffic was complemented by the more routine business and personal traffic typical of the East. Still, before World War I, immigration remained important to CPR's passenger system.

Whatever the purpose of the trip, the fact was that for the late Victorian Canadian, train travel was the norm. Thomas Keefer may have been melodramatic in his predictions of the benefits of the train, but for the public the train did have many emotional associations. It was the train that brought the settler and his family to a new life. It was the train that took relatives to visit one another, and it was the train that provided access to the outside world. The national symbols, in other words, were not remote and distant but a part of the daily rhythm of existence. Even after half a century nothing was more commonly used as a symbol of industrial progress. Only electricity, just beginning to make its appearance at the turn of the century, could even come close.

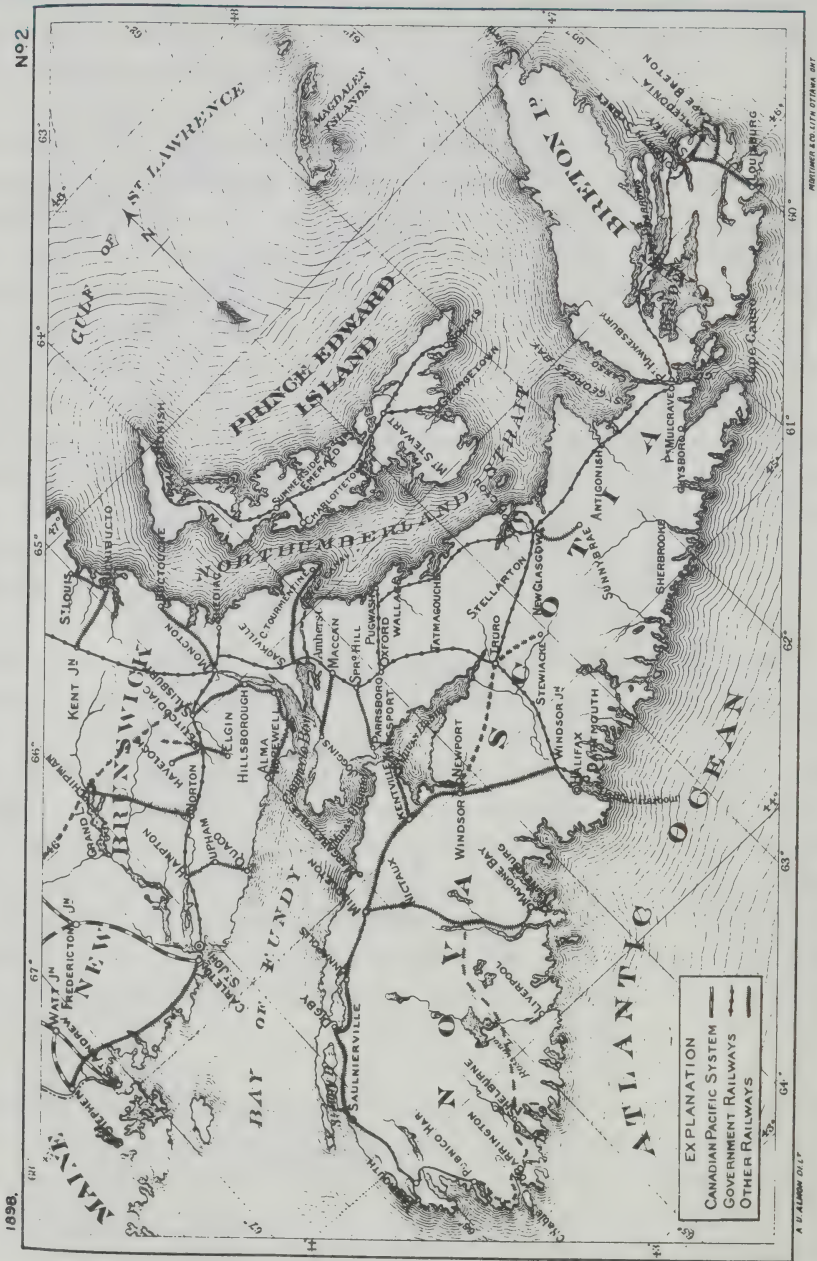
It is possible, therefore, to talk by the late 19th century of a relatively mature railway system. As Figures 3 to 6, which are reproductions of the original maps, indicate, coverage was excellent through central Canada, good in the Maritimes and, though far from satisfactory, increasing yearly in the West. The great projects of Confederation were complete, and the public already viewed their railways as an integral part of both the Empire and the nation. Government relationships to railways were fairly well established.

Figure 3
 THE CANADIAN RAILWAY SYSTEM, 1896



Source: Map reproduced from *Annual Report of the Department of Railways and Canals, July 1, 1897 to June 30, 1898* (Ottawa: Queen's Printer, 1899).

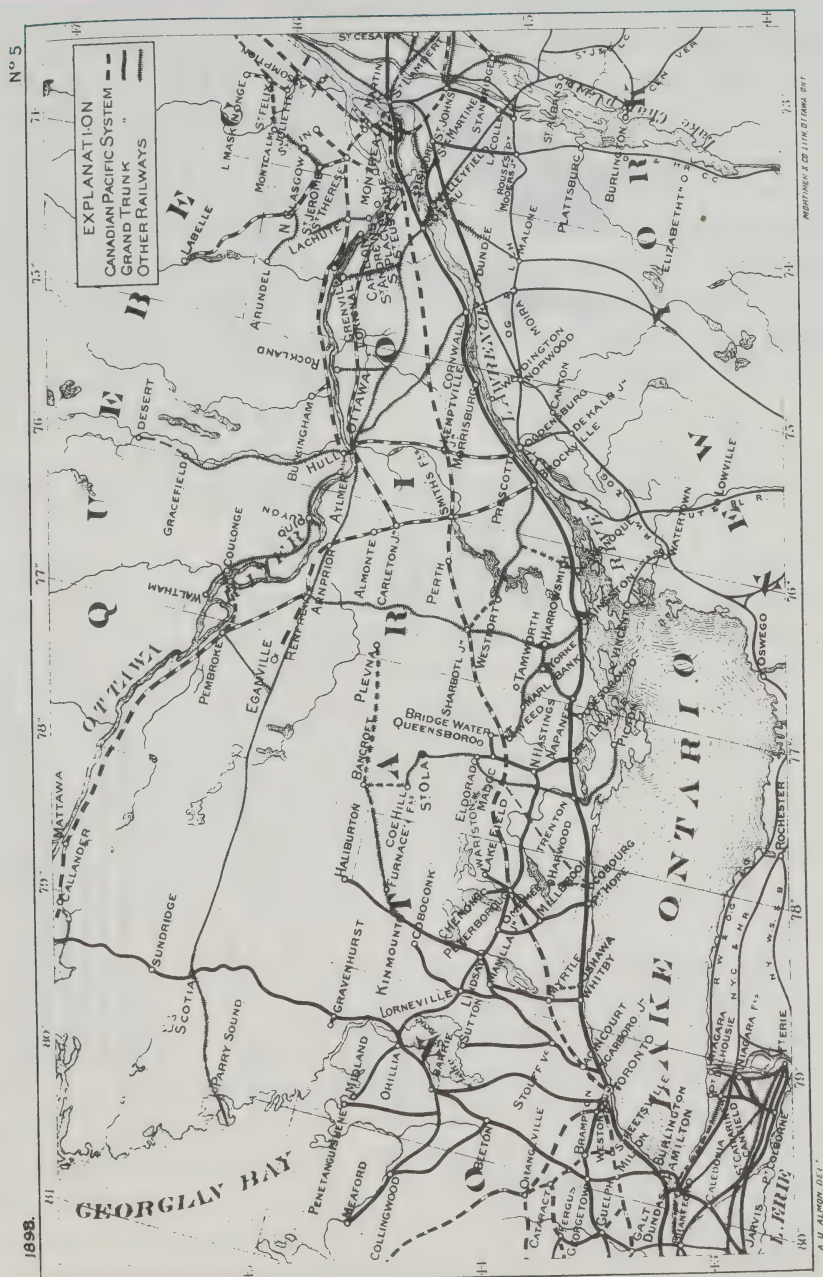
Figure 4
 THE CANADIAN RAILWAY SYSTEM, 1896: DETAILS OF NOVA SCOTIA AND NEW BRUNSWICK



Source: Map reproduced from *Annual Report of the Department of Railways and Canals, July 1, 1897 to June 30, 1898* (Ottawa: Queen's Printer, 1899).

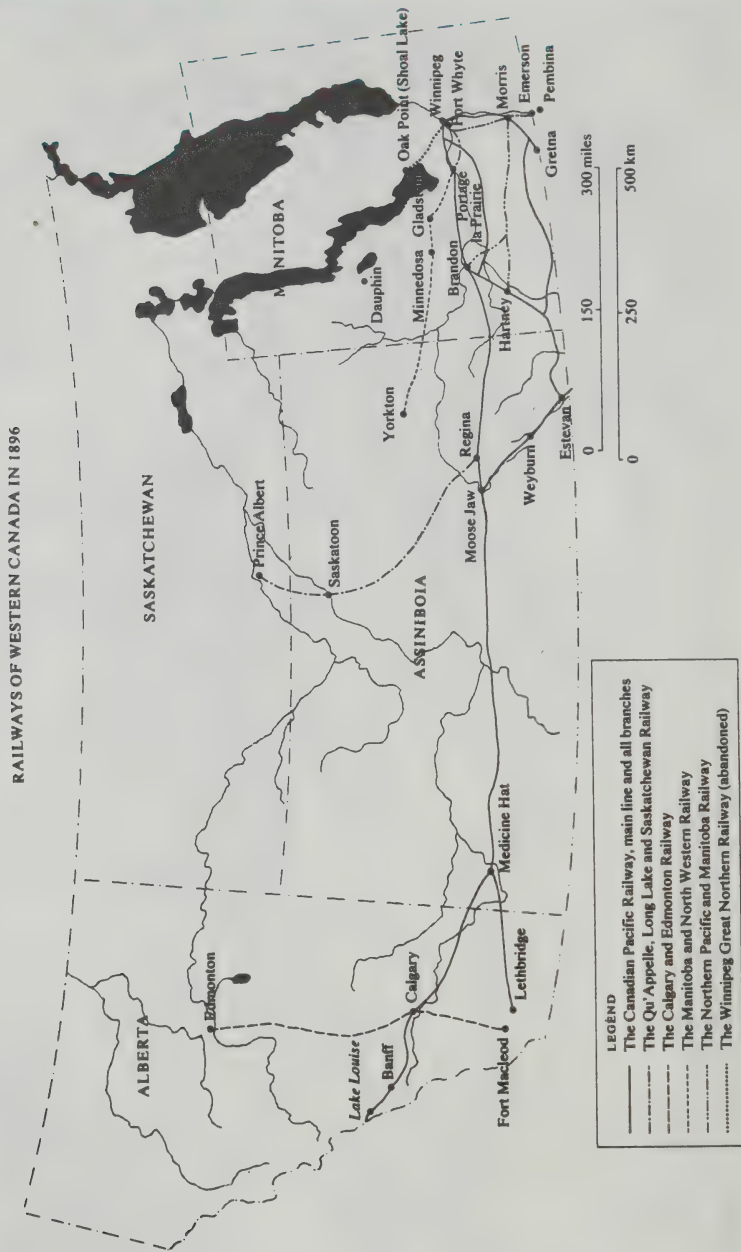
Figure 5

THE CANADIAN RAILWAY SYSTEM, 1896: THE CENTRE OF THE CANADIAN RAILWAY SYSTEM, MONTREAL-HAMILTON



Source: Map reproduced from *Annual Report of the Department of Railways and Canals, July 1, 1897 to June 30, 1898* (Ottawa: Queen's Printer, 1899).

Figure 6
THE CANADIAN RAILWAY SYSTEM, 1896: THE PRAIRIE RAILWAYS



Source: Map reproduced from T. D. Regehr, *The Canadian Northern Railway, Pioneer Road of the Northern Prairies* (Toronto: Macmillan, 1976), p. 22

Governments would aid railway development, often considerably. At the same time government, whether Dominion or provincial, had generally resisted direct ownership. There were the exceptions of the Intercolonial and Prince Edward Island systems. The ongoing operational losses of the two railways, however, generally reinforced rather than challenged the Government's desire to stay away from ownership.

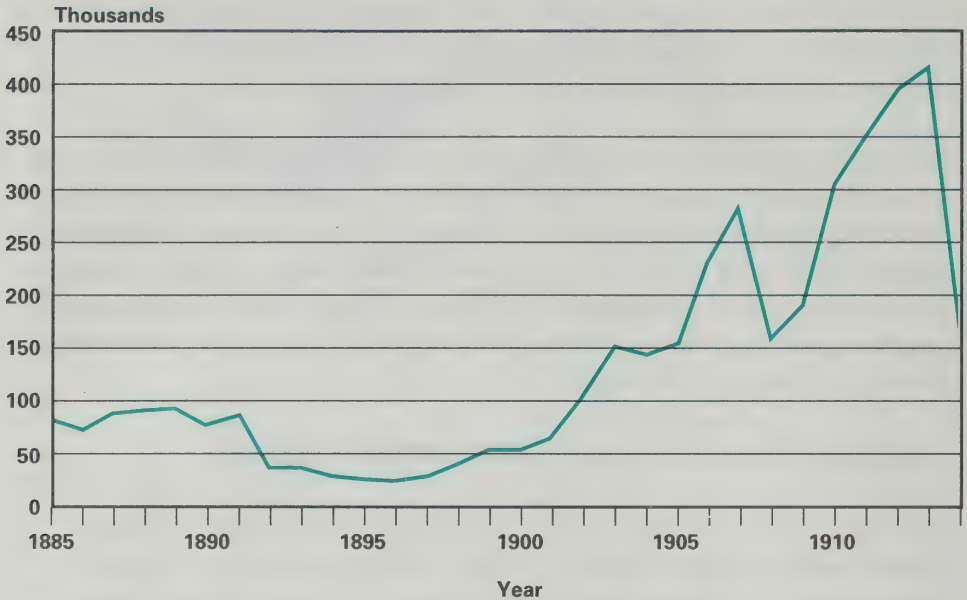
All in all the procedure of land and cash had not worked badly, though the landscape was littered with charters that had been abandoned and offers of land grants that had not been fulfilled. Now, as Canada began the period of greatest expansion in its history, the final great phase of railway expansion was about to begin. Ultimately this expansion would bring the old subsidy program crashing down and forced into creation a network of nationally owned and operated railways.

THE FINAL PHASE OF RAILWAY DEVELOPMENT: 1897–1917

Once again the prize was the West. After 1896, international economic conditions improved. Canada, especially the West, became the destination of choice for thousands upon thousands of immigrants (see Figure 7). Higher wheat prices also provided a rapidly growing export product. Reinforcing these trends was the inflow of British capital to the region. The long-anticipated Western boom was under way, and it continued almost uninterrupted for nearly 15 years. In the 1880s it seemed foolish and daring to talk of one transcontinental railway; now it was feasible to talk of two or even three. The relative stability of the railway system was about to be shaken as the Dominion entered the third and last phase of great railway expansion.

Forces in the West and the East had reasons to want railway expansion. After two decades of CPR monopoly many westerners were disgruntled with high freight rates and with CPR caution in expanding its branch-line network. Two railway entrepreneurs, William Mackenzie and Donald Mann, responded to the disgruntlement and to the profitable charters lying around by constructing a growing pattern of local lines around Winnipeg in the 1890s.⁸³ By the turn of the century their ambitions had grown, and preparations were under way to expand their line, now known as the Northern Railway, eastward to the Lakehead. There was even talk of heading westward toward the Rockies. For Prairie farmers dependent on the CPR such competition was a welcome thing, and they encouraged their legislatures and the company to move forward with the plan.

Figure 7
NUMBER OF IMMIGRANTS TO CANADA BY YEAR, 1885-1914



Source: Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series A350.

In the East it was not regional grievances that were emphasized but the opportunity to make up for a squandered past. The Grand Trunk railway had in the 1870s refused to build an all-Canadian transcontinental line and had therefore been relegated to the sidelines as the CPR became one of the world's most famous railways and was even profitable. The Grand Trunk not only missed the opportunity to become a transcontinental but over the decades since the CPR's completion had been forced to wage an often vicious competition in the East with this upstart rival. Now the Prairie market seemed for the first time capable of supporting a second transcontinental, and the Grand Trunk was determined not to be left behind again.

Both railways followed the established pattern and turned to governments for support. The details of the subsidies, as is so often the case, are extremely complex and only of marginal relevance here. What is important is that a combination of politics, stubbornness and too much optimism meant that instead of choosing between one of the rivals for the new transcontinental

line, Canada got both. Initially the Grand Trunk, along with its newly created subsidiary, the Grand Trunk Pacific, was the favoured child of Prime Minister Wilfrid Laurier's Liberal government. Large subsidies were handed out, and the Government even agreed to construct the most unprofitable portion of the line (north of the Great Lakes) and then turn it over to the completed railway. The Northern Railway refused to back off, however, and used its considerable clout in the West to gain local support, including significant grants and guarantees from provincial legislatures. Then, when Robert Borden's Conservative party took power in 1911 the Northern was able to join the Grand Trunk at the federal trough.⁸⁴

For the next several years the two railways proceeded to build lines across the Prairies along the old Yellowhead route. By 1907 they were in Edmonton, and both were determined to push through lines to the West coast that were, in many cases, practically next to each other. Government largesse, excessive optimism and business egos led to one of the greatest bursts of railway expansion in history. Prairie mileage had tripled to some 11,000 miles (17,700 km) by the eve of World War I. Overall, by 1915, as Michael Bliss has pointed out, Canada was the nation with "more railways per capita than any country in the world."⁸⁵

These railways had been built at considerable expense to the Canadian taxpayer. As a matter of pride and public relations the railways sought to capture the public imagination as the CPR had done. The Northern Railway had the more authentic claim in this regard. With some justification it portrayed itself as the corporate entity that would finally rescue the western Prairies from CPR monopoly. This was the "West's own railway," which by its very existence would force the CPR to be more competitive in rates.⁸⁶ This belief was important both in the political support the Northern gained during its rapid expansion and in the considerable financial support given to it by western provinces like Manitoba. If the CPR capitalized upon the ideal of national unity, the Northern capitalized upon a sense of regional grievance.

The Grand Trunk and Grand Trunk Pacific had a more difficult time attaching themselves to the romance of Canadian railroading. The railway did have a positive image for good service in the East. It was an unknown in the West, however, and the seemingly rich deal it struck with Laurier's government to compete with the Northern did not endear it to Westerners. The railway's managers did take a leaf from the CPR and supported a book extolling the

virtues of the project. Its author looked to the imagery of both the Northern and the Canadian Pacific and made up in rhetoric what the Grand Trunk Pacific lacked in historical tradition. The Grand Trunk Pacific would take on the CPR, which "wielded its power in an autocratic manner." Once completed the "farmer would no longer be at the mercy of a railway octopus." The people who accomplished such works were, of course, heroes:

When the Epic of the railway is written, the men who laid the foundations of the National Transcontinental will loom prominently therein. . . . They were not communicative; hardship, privation, peril, and sensational excitement had been encountered so frequently that they considered such a part of the daily round, and now that they could view them from afar, and at a more distant date, they made light of them. But when I probed beneath the surface, as seated round the blazing fire in the lonely camp in the wilderness I drew them into conversation, and once more threw them back into the days when they were up in the forest toiling mighty hard to find that four-tenths of 1 per cent grade, they grew slightly reminiscent. And what stories they could tell! What thrills they could give!⁸⁷

Despite this author's enthusiasm, the "Epic" was never written. The Grand Trunk railway, though extremely well built, was never able to assert a romantic public image. Even the Northern, though supported by Westerners initially, found its own image floundering within a few years. For what these two railways conjured up in the public mind was not the romance of construction or even the opening of new districts but the sheer cost. Indeed, their experiences shifted the whole image of the railway in Canada generally. Before the new transcontinentals were built the public still revelled in the Victorian imagery of the railways as a symbol of progress and technological triumph. The expectation was that railway construction and expansion were a part of Canada's future. Even among Westerners who distrusted the CPR, for example, there was a continuous clamour that that railway build new branch lines. After the experience of the new transcontinentals the image shifted. Railways became associated with their costs as much as their results. The excesses of the early 20th century meant that henceforth both public and political attitudes to railway development would be much more cautious, even negative.

What happened in the interval is extremely complex. The business affairs of the Grand Trunk and Northern were a massive tangle of land grants, Dominion and provincial government arrangements, bank loans, share issues and holding companies. Amidst all the confusion, however, the story is actually fairly straightforward. First, the railways and the Government overestimated the ability of the nation to support a massively expensive new transcontinental line. Certainly they overestimated its ability to support two such lines. Second, the Government found itself drawn into the same sort of position it had experienced with the canals three quarters of a century before. Loans, subsidies and especially loan guarantees made the Government hostage to the railways. Even before World War I broke out, for example, the Dominion government had acquired some \$40 million in Northern Railway stock, and that was only a fraction of the amount they had loaned out. The railways simply could not be allowed to go under. This was especially true because at least one province (Manitoba) and one bank (Bank of Commerce) were so heavily committed to the railways that their bankruptcy would have followed that of the railways almost immediately. Third, this overextension of resources was made much worse by World War I. A wartime economy pushed prices up rapidly. The costs of supplies, of locomotives and of labour all increased. At the same time the traditional source of foreign capital, the London money market, dried up in the face of wartime demands at home. By 1915 the Northern was effectively bankrupt. The Grand Trunk Pacific was not far behind.⁸⁸

The Government sought answers to the chaos through a royal commission.⁸⁹ The Drayton-Acworth Commission was split ideologically. One faction hoped that a merger of the two railways might allow a continuation of the system in private hands. The other side, and ultimately the Government as well, saw this as unrealistic. The fact was that most of the railway funding, to the tune of a quarter of a billion dollars, was tied to government support.⁹⁰ To put this in perspective, the total revenue of the Canadian government on the eve of the war had been only \$169 million. At the provincial level, British Columbia had guaranteed \$40 million of Northern debt, and Manitoba had guaranteed \$25 million. The annual revenues of the two provinces were, respectively, \$10.2 and \$4.1 million (1913 figures).⁹¹ There was thus no real choice. The federal government decided in 1917 to nationalize the two railways.

The nationalization process was almost as complex as the financial free-wheeling that made it necessary. It was 1923 before the Government railways — the Prince Edward Island Railway, Intercolonial, Grand Trunk and Grand Trunk Pacific, Northern, as well as smaller railways — were formally amalgamated into the Canadian National Railway (CNR) system.⁹² The Canadian government, which had for so long resisted any massive ownership of railways, now possessed nearly 22,000 miles (35,400 km) of track and a massive corporation, spanning a continent and employing nearly 100,000 people. It had also assumed a massive debt and, in a controversial move, put that debt on the books of the new Canadian National corporation. Thus CNR began life with long-term debts of nearly \$2 billion! This made deficits inevitable, and these would be a harsh reminder of the nature of its birth. "The mood of the time in which Canadian National Railways were born was more critical than sentimental; more conscious of the burden that was to be carried than of the ambitions that had been foiled; more censorious of the sins of the fading companies than proud of the achievements they had made."⁹³ The Canadian railway system was, to all intents and purposes, complete.

4. THE "MATURE" RAILWAY SYSTEM AND THE RISE OF THE CAR CULTURE: 1900-1945

Now that the railway system was in place, what did it mean in terms of passenger service? How much was the railway a part of the everyday life of the Canadian? In general terms, the railway network, so overbuilt from a financial point of view, gave the public good access in most settled parts of the nation. In the cities and larger towns, conveyance to the solid brick or stone railway station could often be made by public vehicle. In the larger cities street railways ran regularly along major thoroughfares. The traveller could therefore leave a house in Toronto, Montreal or elsewhere, walk to the corner and catch a street railway to the railroad station. Many passengers took their horse and buggy or, occasionally, went in the new device, the car. Outside the cities, of course, private transportation was all that was available.

In city or country, however, the destination was the train station. For it was there that connections to the outside world were made. In an era before competition, the station combined the role played by air terminal, bus

terminal and, of course, train station today. As the architecture of Union Station in Toronto (completed 1927) indicates, the stations themselves were opulent affairs, reflecting the power and substance of the railway's dominance over the nation's transportation network. The stations were, as the slogan of the time had it, "the crossroads of commerce."

This was more than an idle boast. The major stations were central to business and personal travel. At the major stations scores of trains departed daily.

Once in the station the passenger had access to the continent. Complex routings along different lines and different railways created a wide variety of choices of routes, destinations and levels of comfort. Indeed, the complexity and variety of the turn of the century timetables are more than normally striking for the contrast they present to the current atrophied state of passenger rail service in North America. Also striking and occasionally forgotten amidst the nationalist symbolism was the vast amount of railroad service to the United States. The Canadian Pacific and Grand Trunk both had regular service to all northern U.S. centres. For more distant points they provided transfers and through connections via U.S. lines.⁹⁴

The passenger taking a long trip also had a number of options for sleeping accommodation. The Pullman sleeping car, developed in the United States, was widely used. Canadian Pacific even boasted that it had improved the Pullman and provided comfort superior to that of any U.S. line. The sleeper was also available in a series of classes. First class was a luxurious service, designed to attract the affluent family and business traveller. The tourist sleeper was less ornate but was only about half the cost of first-class accommodation. There was also the colonist sleeper, though the average intercity traveller was unlikely to use or even see this accommodation.⁹⁵ It was really designed for the special immigrant trains that had been so important before World War I.

Also, well before the 1920s, dining cars had become common. Earlier it had been normal for railways to schedule station stops so that passengers might eat — much as bus lines do today. Some of the CPR's first hotels in the Rockies, such as Mount Stephen House, were in fact way stations. The schedules did not leave people much time for a leisurely lunch. Twenty-minute to half-hour stops were considered adequate.⁹⁶ Such tight timing led to discomfort for passengers and undoubtedly delays for the crew. By the

20th century, however, the well-appointed dining car was one of the features expected by first-class passengers. These cars rarely made money for the railways, but they were a necessary service if railways were to compete. On longer trips and the most important lines, therefore, dining service was regularly available.

There were some alternatives to trains. On the Great Lakes and along the British Columbia coast, travellers could take steamships, many of them owned by CPR and integrated into train schedules.⁹⁷ For other trips people could simply travel by horseback, carriage or car. Such individual travel is impossible to measure, but over long distances it was certainly an inferior means of getting about. Nothing, however, could compete with the train, which completely dominated both commuter traffic and the long haul. In 1920 there were six train trips taken for every man, woman and child in Canada.⁹⁸ Fifty years later it was less than one — and the relative cost of train tickets had decreased in the interval. In the years between the turn of the century and the Great Depression, the normal business or personal traveller almost certainly took the train to travel between cities.

To maintain their dominant position after the war, however, the railways did have to develop new strategies. Before World War I, as has been discussed, two types of passenger predominated: the local commuter (mainly in Eastern Canada) and the immigrant. Though the train was a part of the rhythm of life, it was also a means to other unrelated activities — settlement, business or visits. By the 1920s, however, there were growing concerns about the future of both kinds of traffic. Most of that concern was focussed on the apparent end of massive immigration to Canada. Before World War I, immigration had reached historic highs, averaging more than a quarter of a million people a year. Naturally, the war brought a precipitous decline: immigration reached a low of 36,000 in 1915. More significantly, the recovery after the war was incomplete. The days of the “last best West” as a major destination for immigrants was over. Canada’s primary agriculture frontier was now taken. This, combined with more restrictive rules, meant that half as many immigrants came to Canada in the 1920s as before the war. With a larger percentage of these staying in the urban areas of the East, immigrants could not provide the staple passenger traffic for the railways that they once had.⁹⁹ If the immigrants weren’t coming in as great numbers, however, the railways hoped that the tourists would. In the 1920s the railways put a major effort into attracting the lucrative tourist dollar.

Tourism was not new to the railways. From the beginning the Canadian Pacific had made efforts to draw tourists to the West. Passenger agents in London and the United States, although concerned primarily with immigrants, distributed promotional pamphlets emphasizing the grandeur of the scenery and the comfort of the railway.¹⁰⁰ William Van Horne himself liked to make up advertising slogans, though they were noted more for their unusual nature than their staying power. The railways also gave free passage or other support to well-heeled travellers who intended to publish their experiences.

From the beginning the railways had tried to sell more than mere railway fares to tourists. As land sales attracted immigrants, so an integrated CPR package could attract tourists. By 1886, one year after the railway opened, there were three hotel dining stops in the mountains.¹⁰¹ The next year, with the active encouragement of the CPR, Banff National Park was created by the Dominion government. The CPR quickly built a hotel on the site, and in 1888 the first version of the Banff Springs Hotel opened for business. By 1891 it was receiving more than 3,000 guests a year, almost all arriving by railway. Over the next decades the network was improved and extended. The "Chalet" Lake Louise (1890)¹⁰² and Hotel Vancouver (1887) reinforced the network of hotels and emphasized the fact that tourism for the CPR increasingly focussed on the mountains. By the turn of the century the fare-paying passenger had an annotated timetable to describe local sites, several luxurious hotels at which to stop over, and possible side trips and excursions at various points along the way. There were also 115 sleeping cars to cater to the luxury, largely tourist, trade.¹⁰³

All of this paled next to the efforts of the 1920s. The two railway companies made massive investments in the tourist industry. As one historian has said, "the hotel and resort war that took place in the later twenties cannot be explained by rational projections of expected profits."¹⁰⁴ It can be understood, however, by realizing that for the first time two complete transcontinental systems sought supremacy. Each resented the other, and both were led by determined, energetic individuals. Accordingly, millions of dollars were spent on constructing or renovating hotels still familiar today: the Bessborough in Saskatoon, the Nova Scotian in Halifax, the Chateau Frontenac in Quebec City and the Chateau Laurier in Ottawa. And in Toronto, Canadian Pacific built what was supposedly the largest hotel in the British Empire, the Royal York.¹⁰⁵ At \$16.5 million it cost more than the annual revenue of six of Canada's provinces! Of course these hotels were for business

travellers as well as tourists. They were also a part of the railway's image, an image that would be increasingly difficult to maintain when the Great Depression hit within a few years.

Despite all these expenditures, the railways continued to focus their efforts for the tourist dollar in the mountains. By the 1920s the earlier facilities at Banff and Lake Louise had grown in size and splendour. To the north the Canadian National Railway built Jasper Park Lodge (1922). Along the CPR route through Banff, detailed guidebooks were available. These not only told you of the sites to be seen but all the activities that could be arranged as a part of your trip. Some of this was tremendously organized. It was possible, for example, to detrain at Banff, meet a prearranged guide, tour Banff, travel to Lake Louise ("a fire proof and modern hotel with accommodation for seven hundred guests"), spend the night there, get up in the morning for a motor trip to Emerald Lake and then go on to meet the train. The cost was \$18.50.¹⁰⁶ Side trips were also possible at various sites in British Columbia.

Around the major cities of the country, railways provided another type of tourist service. From the 1890s on, the growing urban population of Canada, especially in Ontario and Quebec, had begun the ritual of escape to cottage country. Private cottages or resorts dotted along the Laurentian and Muskoka lakes created a lucrative market for trains. Even by the early 1920s cars were still relatively rare and a 100-mile (160 km) trip enough of an adventure that the trains captured a good deal of the trade. By the end of the decade, however, the car was beginning to have an effect on this trade.

Tourism was important not just for the railway revenues but for the way in which it further integrated the railways into the life of Canadians. Earlier sightseeing and adventuring across the wild West had now been replaced by a more sedate, widespread and family-oriented tourist trade. The grand railway hotels added to the glamour and romance of the railway system. The railways accordingly became associated with both the Canadian landscape and Canadian social activities — whether dining in the major urban hotels or backpacking in the Rockies. All of this added to the belief in the importance of the railway passenger system to the country as a whole.

Tourism and the increasing travel of the urban business class made the 1920s generally good years for the two major railway systems. Freight revenue was high, but so too was passenger traffic. In 1920 more than 50 million

passengers were carried on the railway system, an all-time high. Throughout the 1920s numbers well in excess of 40 million were common. Moreover, revenue for passenger traffic per train-mile increased in real terms throughout the 1920s. Overall, passenger revenue was an important part of the earnings of both CPR and CNR. In 1925, for example, CNR earned 19 percent of all revenue from passenger fares.¹⁰⁷ In contrast, 40 years later it would be 7 percent.¹⁰⁸

The importance of these revenues was reflected in the service. Both major railways had more extensive passenger schedules in the 1920s than they had before or would after. The tourist who wanted to go westward from Calgary to Vancouver, for example, had a choice of four trains daily.¹⁰⁹ By the end of the decade the CNR had inaugurated a train that travelled from Toronto to Montreal in six hours. Even on the less popular runs the service was impressively complete and efficient. Many trains stopped at literally hundreds of communities while wending their way from one major centre to the next. Moreover, though major new routes were not opened in the 1920s, there had been continued construction of branch lines, especially on the Prairies. Thus despite of some closures and rationalizations resulting from the collapse of the Grand Trunk and the Northern, railway mileage in Canada increased by some 3,000 miles (4,800 km) in the course of the decade.¹¹⁰ By that point most communities of any size in Canada outside of the North had access to a railway. At the end of the decade the train remained the common means of travel and a part of the ordinary daily routine and the special occasions of Canadians' lives. The railways remained king of passenger conveyance and passengers remained important to the railways.

THE RISE OF THE CAR CULTURE AND ITS CONSEQUENCES: 1900–1929

Even as the railway system matured a new technology was rising to challenge the dominance of the railway. In the 30 years after the turn of the century the car would revolutionize Canadian life. Government attention would turn increasingly from the overbuilt and expensive railway system to roads. The public itself, though far from abandoning the railway, found the car's flexibility and privacy tempting. Though another generation would pass before the car overwhelmed the railway as a vehicle for passenger service, the changes that took place within a generation were, nonetheless, spectacular.

At the turn of the century the car was still a toy, owned by only a handful of Canadians. Indeed, the first Canadian car would not be built until the next year. None of the legal structures, roads or governmental financial systems were designed with any thought of the car. There were no licensing provisions. In fact there were no rules at all except those that could easily be borrowed from statutes aimed at horse and carriage. In one generation all this changed. After the turn of the century the car became an increasingly common sight on Canadian streets. By 1914 there were more than 50,000 vehicles registered in Canada.¹¹¹ In the 1920s ownership exploded, turning the automobile into an everyday form of transportation for millions of Canadians. By the end of the decade Canada had the second highest per capita ownership of motor vehicles in the world, more than double the rate of the United Kingdom or France. The legal, cultural, governmental and fiscal structures of the nation had had to adapt to this new technology. Along the way the basic form of the modern Canadian road system and road policies emerged. Sometime in the period between 1900 and the later 1920s the car became a fundamental part of Canadian transportation.

There was no single inventor of the automobile. Rather, from the time of the industrial revolution people had tried to harness motorized power to the carriage. It was not until the 1890s, though, that the car was distinguishable as a new vehicle, though the technology remained primitive. Shortly after the turn of the century cars were numerous enough that laws began to appear on the books to deal with them. This was especially necessary in the United States, which was rapidly becoming the world's most motorized society. As early as 1901 New York State passed driving regulations and laws, and by 1905 most states had speed limits.¹¹²

Canada was not far behind. Like the United States it had a prosperous population with large disposable incomes. Canadians also had obvious reasons for being interested in improvements in transportation, given the size of the country. Also, except for a few areas, Canada was thinly settled and that meant that fixed forms of transportation like the railway could not possibly serve every community. A land vehicle freed from the track but able to achieve relatively high speeds possessed great potential for a geography like Canada's. Thus, living next to the United States, prosperous in its own right and with a real need for transportation innovation, Canada soon followed American footsteps into the automobile market.

The first Canadian car was built in 1901 by the T. A. Russell Company. Before long other entrepreneurs and innovative carriage makers followed suit. A competitive and as yet only loosely formed auto industry soon developed, mainly in Ontario. Ontario was also the province where the car was first seen with any regularity. In 1907, the first year for which statistics of car ownership are available, Ontario accounted for more than 70 percent of cars nationwide. Outside of the Windsor–Montreal corridor and the Vancouver–Victoria area, car ownership was still extremely rare. Only in this corridor were there sufficient wealth, population and, not incidentally, passable roads to encourage any sort of car ownership.

Even at this early date, however, the rise of the car had forced governments to respond. Ontario passed a registration provision in 1904, and within a few years other provinces followed suit.¹¹³ New technology also brought new rules of the road. The car operated at high speeds compared with horse and carriage, and this created concerns about safety. Accordingly, it was not long before traffic violations and traffic accidents became an item of concern. Speed limits (usually 15 to 25 miles per hour (24 to 40 km/h)), regulatory codes and lists of fines and punishments became necessary. Only in Prince Edward Island was all this legislative activity muted. That province resolved the whole issue by prohibiting motor vehicles outright. Even when, in 1913, it decided to allow motor vehicles the province was hardly enthusiastic. Driving was banned Tuesday, Friday, Saturday and Sunday. Perhaps it is not surprising that the province had a mere 31 cars by the beginning of World War I.¹¹⁴

As society came to grips with the new technology, auto makers continued to improve their craft. Companies collapsed and new ones arose. Giants like Ford and Buick and Oldsmobile began to take a larger share of the market. The open car gave way increasingly to the closed one — a very practical development given Canadian winters. Air-cooled engines gave way to water-cooled ones. Steam engines yielded to gasoline engines. Horsepower increased, and at the top end of the market considerable elegance was possible. There was also tremendous development of ancillary industries. Parts became increasingly available, as did gasoline and service.

Most important of all, however, was the arrival of the relatively cheap and reliable car. In 1905 the average price paid for a car was in the neighbourhood of \$1,784, or about 18-months' salary for a middle-class white-collar

worker.¹¹⁵ Then, in one of the events that has become a legendary myth of industrial history, Henry Ford revolutionized production and design to produce the famous Model-T in 1908. Powered by a four-cylinder, 20-horsepower gasoline engine, the Model-T provided reliable transportation for \$850. By the early 1920s up to a million Model-Ts were being sold every year in North America. Other companies quickly adjusted or, if they did not, sank into bankruptcy. By the 1920s several cars, such as Chevrolet, could reasonably claim price parity with the Model-T and superiority in design and comfort. The Model-T and its direct competitors made the car available to the middle class.¹¹⁶

These innovations accelerated the rate at which car ownership increased in Canada. As Table 1 indicates, the number of Canadian-owned vehicles in 1907 (2,130) had increased tenfold to 21,519 by 1911, and this number had more than tripled (69,598) by the beginning of World War I. Also, ownership had spread more evenly through Canada. Ontario drivers then owned less than a third of Canadian vehicles, and even Prince Edward Island had eased its anti-car position by World War I. As we saw above, it possessed 31 vehicles. The car had become a national institution.¹¹⁷

Table 1
AUTOMOBILE REGISTRATIONS, 1907-1929

Year	Number	Year	Number	Year	Number
1907	2,130	1916	123,464	1926	728,005
1908	3,033	1917	197,799	1927	836,794
1909	4,763	1918	275,746	1928	945,672
1910	8,967	1919	341,316	1929	1,070,664
1911	21,519	1920	407,064		
1912	34,136	1921	465,378		
1913	50,558	1923	513,821		
1914	69,598	1924	586,850		
1915	89,944	1925	652,121		

Source: Statistics Canada (formerly Dominion Bureau of Statistics), *Canada Year Book*, 1933, p. 686.

The real boom was yet to come. In 1918 the Mayor of Hamilton, surveying the likely demands of the post-war world, warned "This is not the time of railroads, but this is the time of good roads, and it is the time of automobiles. The traffic conditions have absolutely changed during the past few years."¹¹⁸ Despite the recent collapse of the Grand Trunk and Northern it was far too early to sing the death knell of the Canadian railway. Nevertheless, the Mayor

had recognized an important point. The post-war world was car crazy. Even in 1921, when a vicious slump devastated many sectors of the Canadian economy, both automobile registrations and production continued upward! The Mayor was right. The time of the car had arrived.

James Flink, in his assessment of American automobile development, uses the term “car culture” to emphasize the pervasive influence the automobile has had on modern life. The car, Flink argues, “more profoundly influenced twentieth-century American historical development than the collectivity of reforms emerging from the so-called Progressive Era and the New Deal combined.”¹¹⁹ At some point the car became a sufficiently powerful presence that society began to revolve around it in myriad ways — from town planning to advertising to the economy. Thereafter nothing would ever be the same, and the arrival of the car culture, as much as World War I might be taken as the true end of the 19th century.

For Canada the car culture arrived in the 1920s. Whether judged by the diffusion of the technology, form of production or supporting infrastructure, the 1920s was a watershed decade for the car in Canada. The number of registrations rose to more than one million, and the number of persons per car dropped from 30 to 10. Throughout the decade registrations increased at an average rate of well over 10 percent a year. At the same time the horse declined in numbers for the first time since statistics had been kept. It would not recover.¹²⁰ On the production side, automobile manufacturing underwent a major series of consolidations, including the collapse of the last independent Canadian manufacturers by mid-decade.¹²¹ It was also in these years that many of the modern distribution, marketing and advertising techniques associated with the car came into being. The development of the car culture also saw the formation of a more effective infrastructure. Automobile clubs, which had been irregular and largely local before the war, were expanded and organized. Likewise, anomalies in driving regulations (such as left-side driving) disappeared. Only one out of 10 Canadians may have owned a car, but the values of the automobile, its appeal to people and therefore its demands upon governments were far greater. The car culture had become a way of life.

The modern image of the car is ambivalent. Traditional images of romance compete with fears of pollution and safety. So it has been from the beginning. Carriage drivers protested against the fright to their horses caused by the

noise and speed of cars. The absence of driving tests and the basically dangerous nature of the vehicle quickly made traffic accidents an issue. Given these decidedly mixed reviews, the striking thing is the hold the car had on imaginations, even by the 1920s. This was not just a new appliance such as, say, the electric range, but a new experience, and the imagery of that experience must be understood if the place of the car in 20th century Canada is to be understood.

There is a fact that precedes the mythology, however. The automobile was an improvement over the horse, not just in personal but in social terms. The concept of *The Horseless Age*, as one American automobile journal was entitled, was appealing for many reasons. When the car was introduced it appeared to possess considerable sanitary and social advantages over the horse. The horse was an animal and that created obvious problems. In New York City, for example, about 1,250 tons (1,130 tonnes) of horse manure was removed each year. There were also thousands of horses found abandoned or dead by city officials annually. Both the refuse and the carcasses created horrendous odour and health problems. Horses also got ill, often with infectious diseases that threatened surrounding animal life and caused suffering for the animal. At least when a car broke down it did so in mechanical solitude. In sum then, the automobile was "cleaner, safer, more reliable and more economical than the horse."¹²²

The image of the car has never rested on such mundane matters as health and economy, however. Practically from the beginning, enthusiasts attached a whole host of values to the car. Before long it had developed a mystique of its own. Sometimes this mystique could be expressed in terms as romantic and all-encompassing as those used by early promoters of the railroad. Chevrolet advertised its cars under the banner headline, "What Transportation Means to Civilization." Civilization, it determined, was the result of the interchange of individual thought, "and that thought could only be exchanged if transportation was available." The car was the superior form of transportation because "it is not limited to mass movements nor [sic] fixed routes" but could penetrate everywhere. It was, accordingly, the great civilizer of the modern age.¹²³ This was a common theme in the advertising, but it was far from restricted to advertising. W. H. Dandurand of Montreal, one of the foremost early car enthusiasts of Canada, sounded reminiscent of 19th century railway philosopher Thomas Keefer as he described what improvement of a highway did for the region south of Montreal: "The

schools are kept in good shape, the children are better dressed, the little girls wear new dresses and all the children wear boots. . . . Last year we noticed that they had flowers on their lawns and that they had real lawns on which they used lawn mowers and kept the grass nicely cut.”¹²⁴ All this because the car could now readily traverse the area.

Such hyperbolic rhetoric points to an important theme of the age. This was a time of rapid urbanization. People feared that the movement to the city would undermine a moral fabric that, it was believed, rested in the countryside. The car provided a partial remedy to the problem. Rural families might be less tempted to throw their lot in with the cities if they were less isolated. The car provided an easy access from farm to city. On the other side, those who had already cast their lot with the city might at least escape the oppressive weight of concrete and brick to return to the countryside. Henry Ford summed it up with the comment “We shall solve the city problem by leaving the city.”¹²⁵ The President of the University of Toronto, Robert Falconer, said much the same thing in 1915: “Is it not going to be possible to get people from the cities out into the country? Is it not possible to relieve the pressures of the cities?”¹²⁶

These broad social values only indirectly translated to a personal appeal to the would-be car buyer and driver. In selling the car, advertisers used themes of status, taste, refinement and culture. Many of the themes would still be familiar today. A 1929 advertisement in *Canadian Motorist*, for example, portrayed a chic, modern woman standing beside the newest Hupmobile. “Sports ensemble by Regny. . . . Car by Hupmobile,” ran the ad. An Essex Super Six sought to convince prospective buyers that they were conforming to popular taste, referring to “this country and world-wide conviction” about the quality of the car. Yet another ad — this one from an American publication — reassured potential buyers that to consider the 1921 Liberty put them among those “whose taste and means incline them to better things.”¹²⁷ Such imagery was part of the new advertising culture that had developed in the years since the turn of the century. Indeed, the car and modern advertising grew up together, each reinforcing the other’s hold over the public. The concept of being smart and elegant was used to sell everything from furniture to cigarettes. There were also aspects of the car that, though played upon by advertisers, were closer to the real mystique of the vehicle.

Central was the idea of independence. This was your personal vehicle that could take you where you wanted when you wanted. In dozens of advertisements and articles the theme of independence was promoted in myriad ways. To a surprising degree, women were depicted as the driver of the automobile. The message conveyed by those ads seemed to emphasize several things, all related to independence. The ads implied that it was easy to operate the car and that the modern woman need not fear this new technology. Any woman could do it. This was no small matter to a generation of women who, born and raised in cities, had increasingly less knowledge and skill in horsemanship. It was also a message not lost on the urban man who, though less willing to admit such things, was himself less skilled with the horse than his forefathers had been. "It's a man's car but it's a woman's car in the sense that it's so easy to operate," said an Overland ad in the early 1920s.¹²⁸ The image of the woman driver also linked the car to the modernity of the 1920s. The old Victorian days were gone. A woman who had access to a car wasn't dependent on some man but was free to go about her business. So efficient was the car, moreover, that family duties need not be forsaken in doing so. There was adventure to be had out there beyond the world of the family and the confines of the home. The car provided the means to discover it.¹²⁹

The advertisements emphasizing independence were so powerful because they reflected a genuine public view of the car. Articles on motoring in the early years focussed less on mechanical attributes and speed than today's magazines and more on the romance and adventure of touring. The car was a vehicle that freed the individual from the fixed lines of the railroad and opened up whole new areas of countryside. Driving was still sufficiently new and the countryside sufficiently uncluttered by other cars that there was a real sense of adventure to motoring beyond the city. People, in the words of one article, became "Knights of the Road" heading out to see what could be seen. Trips to the end of the road in Cape Breton or into the Rocky mountains were portrayed in 1929 in terms that are associated more with modern backpacking than with motoring in the 1990s.¹³⁰ The car was flexibility, independence and adventure rolled into one. A 1929 advertisement entitled "Beyond the Barriers of Everyday" caught the mood perfectly. The picture was of a family having a picnic in a wild, wind-swept setting reminiscent of a Group of Seven location. The father stands in the centre of the picture as if guarding the situation while a very lavish picnic is spread about. "Break down the barriers of everyday," the ad read. "There's happiness

ahead. . . . The outstanding Chevrolet is the magic carpet that lifts us far beyond the barriers of everyday life. The joys that lie round the turn of the road or over the hill. . . . You too, can seek and find them in your Chevrolet.”¹³¹

The growth and development of Canadian roads must be set against the popularity of the automobile. Cars need roads, and it more or less followed from this that car enthusiasts were road enthusiasts. As the car became a part of Canadian culture the demand for roads grew. The translation of this demand into policy during the first generation of car ownership was extremely important. For by the late 1920s the whole complex set of lobby groups, policies and taxes that shaped roads in this country was in place. So too, to a surprising degree, was the modern Canadian road system, at least in embryonic form.

Support for improved roads predated the car. By the late 19th century both urban bicyclists (a new craze) and rural agriculturists were dissatisfied with the state of Canadian roads. Moreover, this was an age of reform — inter-locked groups of social reformers were advocating everything from temperance to better sewers. Amidst the reformism there was, as well, social insecurity.¹³² Farmers in central Canada in particular were nervous about the future as an increasing number of their children headed into the cities. Industry seemed poised to replace farming as the most important economic sector in Canada. Rural Canadians were an important force, however, constituting some 70 percent of all Canadians in 1891. They were determined to preserve their position in the community. New parties sprang into existence. Various campaigns sought to bring the benefits of the city to the countryside, hoping thereby to remove the temptation of the city. Over the next few years rural telephone systems and rural electrification would become major issues. Improved roads fit nicely into this program. Improvements would presumably reduce transportation costs and thus improve the farmer’s standard of living. They would also reduce the isolation of the farm family and, it was hoped, lessen the attraction of the city.

Still, compared with the massive support of such movements as temperance, a small meeting at the Canadian Institute in Toronto on February 9, 1894 did not seem very significant. Indeed, Andrew Pattullo of Oxford County opened the meeting with the comment that “the attendance will not be very large this morning,” and he apologized for the absence of any important political figures. It was in this unlikely atmosphere that the Ontario Good

Roads Association was formed, and the date is as good as any to mark the turn-around of the Canadian road system.¹³³ For despite Pattullo's deprecatory comments, the good roads movement in Ontario received a significant amount of attention. Cabinet ministers and senior government officials regularly attended. Before long good roads associations sprang up in other provinces. By 1914 F. H. Dandurand of Montreal had successfully formed a Canadian Good Roads Association. In 1917 it was chartered by the Dominion Parliament.¹³⁴

The good roads movement was begun by farmers. Initially, farm associations and ministers of Agriculture showed the greatest interest in it. It was the rise of the car culture, however, that turned the good roads movement into a national cause. Moreover, the movement, though very much the outgrowth of the era's reformism, was unique in the way it blended politics and science. The annual meetings were combinations of technical discussions of road building, discussions of motoring activities and political rallies designed to impress government and public both with the need for better roads. The associations also became significant political forces. From the beginning they attracted the attention of provincial governments. In 1915 (the first year the national proceedings were published) a combined Dominion-Ontario meeting in Toronto showed that the elite members of society had recognized the importance of good roads. The Mayor, the President of the University of Toronto, provincial civil servants and government ministers all attended. It was thus no exaggeration when the Ontario Minister of Public Works concluded that "You have laid the foundation by creating a public opinion" on the matter of roads and that as a result the principles are no longer debated. It is a question of "devising ways and means to meet the different situations."¹³⁵

In 1915 the car was not yet a mass phenomenon. After World War I the car culture took hold, and so too did the Canadian Good Roads Association (CGRA). In 1920 the CGRA held its annual meeting in Winnipeg, the first one held outside of central Canada. Through the 1920s newspapers carried extensive coverage, and the old central Canadian organization was increasingly a national one. In 1925 the Governor General, Lord Byng, attended its annual banquet, and national as well as provincial figures clamoured to be associated with the organization. The attraction of the movement reflected both the public mood and the prestige of a lobby association that was dedicated to the scientific improvement of the Canadian road system.¹³⁶

Paralleling the good roads movement were the provincially based motor associations that would eventually evolve into the Canadian Motor League. A U.S.-inspired network of non-elite associations of local automobile clubs dedicated to the promotion and assistance of “automobility” soon took root. The development of these associations varied considerably from province to province. By World War I the Ontario Motor League was a well-established body with connections to both the Ontario Good Roads Movement and the government. In other provinces a provincial body was still well in the future. It was 1926, for example, before the Alberta Automobile Association came into being as a result of the merger of local clubs.¹³⁷

Like the good roads associations, the motor associations became lobby groups. As early as 1915 the Ontario Motor League was instrumental in arranging a reciprocal licence agreement between Ontario and New York State. Previously drivers had to buy a New York licence plate before they could cross the border.¹³⁸ Interprovincial and international reciprocity soon became the norm, thus extending the horizons of the car driver and the potential of the tourist industry. Other campaigns followed, including one to get all Canadian provinces to have motorists drive on the right. As the automobile numbers increased the motor associations also campaigned for more traffic safety. The motor associations also collaborated with and supported the good roads associations that had grown up in various provinces.¹³⁹

There were differences in the evolution of the two types of associations, however. The good roads associations quickly developed a technical interest, and the problems of road construction as much as government policy became their focus. Also, the various provincial good roads associations were, by the 1920s, overshadowed by the national body. In contrast the automobile associations had always had a “facilitation” role. They were the clubs of the drivers, and therefore travel guides, insurance schemes, international driving permits and club associations dominated their non-lobbying interests.¹⁴⁰ They were, overall, definitely non-technical in orientation. Also, in contrast to the good roads movement, the motor associations remained provincially based. Though the Canadian Automobile Association (CAA) was formed in 1915, it was merely a paper organization until 1957. Even today the CAA is small compared with the huge provincial counterparts that underpin it.

Several forces thus came together to shape road policy in the first 30 years of the 20th century. The primal force was the new technology of the car. Both as a form of practical transportation and as a cult of adventure and

freedom, it created a totally unprecedented demand for good roads. Already in place as that demand grew, however, were a series of provincial organizations and then a national one dedicated to the improvement of roads, connected to the engineering and political elites of the day and seemingly well received by press and public. Then, parallel with the growth of the car came the car club, lobbying in the name of the driver and linked to the good roads associations, sometimes formally yet always informally.

These developments meant that government neglect of roads soon changed. Even before the car really took hold, the good roads movement and the demands of modern industrial society began to affect policy. In Quebec the provincial government moved through the Department of Agriculture in 1897 to give subsidies to municipalities for road improvement. In New Brunswick where intermunicipal roads were under direct provincial control, planning and funding improved. In Ontario a "Provincial Instructor in Road Making" was appointed in 1896, but the most important step came in 1901 when the *Highway Improvement Act* was passed. This provided a subsidy from the province to county road operations and marked the beginning of a long-term trend toward centralization in road construction. In Quebec, Acts in 1907, 1911 and 1912 regularized the principle of a provincial presence in both the funding and construction of roads.¹⁴¹

Centralization was necessary, for the car made the jurisdictional boundaries of the 19th century obsolete. When roads were relatively marginal feeder systems, it was enough for local units to take responsibility for them and to use the inefficient statute labour system to maintain them. Once the car arrived this had to change. For one thing, the longer distances now travelled on land meant that the burden on roads in an area might come from residents miles away. For another, the growing importance of the road system demanded some consistency of funding and construction. As early as World War I, for example, the City of Toronto realized that there was little point in improving its own roads if, once the traveller hit the city limits, travel ground to a halt. It therefore convinced the Ontario government to create the York Highways Commission to develop and maintain roads for 15 miles (24 km) around the city.¹⁴²

The inappropriateness of local government being responsible for road policy after the advent of the car was recognized across the country. The next decades saw a trend toward centralization of road responsibilities. By World

War I most provinces had a distinct roads department or a division within departments of public works. The previously purely local system was complicated by a series of statutes and policies that provided subsidies to local units of government and saw the provinces directly involved in road construction. There was even federal involvement. In 1911 Robert Borden's Conservatives came to power. Borden pledged to take a more active role in facing modern problems than had his laissez-faire predecessor, Wilfrid Laurier. One of the first steps the Government took was to introduce a bill to provide federal funds to assist in the development of a national trans-Canada highway. The bill was defeated by a hostile Senate, but at the end of World War I the Government reintroduced it in slightly different form and passed it as the *Canada Highways Act*.¹⁴³ The involvement of the federal government, along with the growing role of the provinces, indicates just how far roads had come from being a purely local matter.

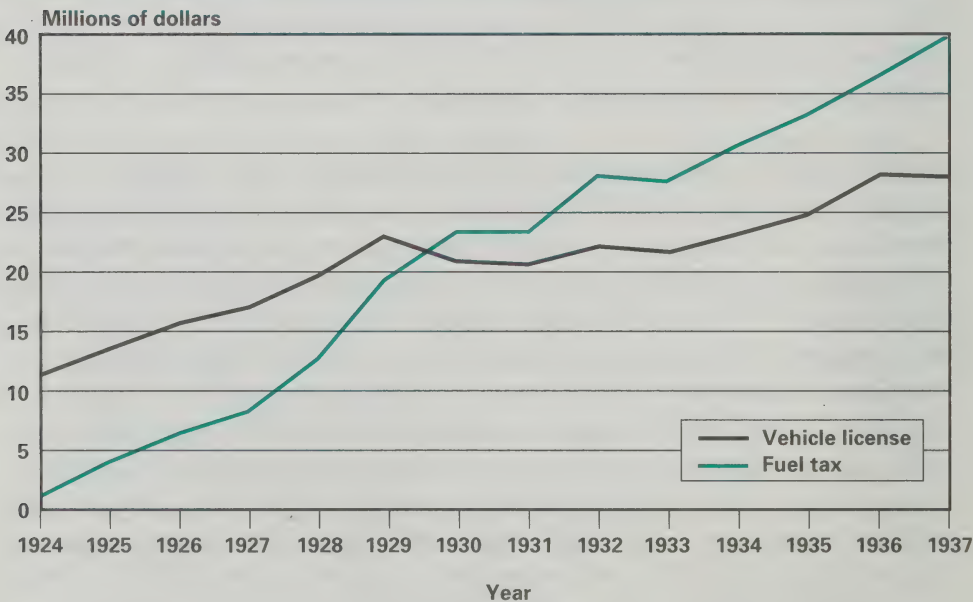
The emerging programs of road subsidy and construction at the provincial level were extremely expensive. Fortunately, the car, which was what created the burden, also provided the solution. Taking their precedent from the United States, the provinces began to impose taxes on cars to cover the costs involved in supporting them. Motor vehicle licensing provided an early source of revenue. Ontario began the practice in 1903, registering all of 220 vehicles that year.¹⁴⁴ Before World War I the registration or licence fee was also adopted by other provinces.

By far the most lucrative source of revenue, however, was inspired by a number of Western states shortly after the war. Beginning in 1919 Oregon, New Mexico and Colorado taxed fuel to raise money specifically for road construction. By 1925 the tax had proven so lucrative that some 44 states had adopted it.¹⁴⁵ Not surprisingly, the provinces were quick to notice the tax. They were facing unprecedented costs as the demands for roads grew. The *Canada Highways Act* only made the pressure greater: it had provisions for Government funding on a cost-sharing basis. If the provinces were going to take full advantage of this federal largesse, they were also going to have to pour large amounts of money into the road system. The fuel tax had the dual attraction of being lucrative and focussed on the very group that would most benefit from the new roads — the car user. In 1922, Alberta was the first province to implement a fuel tax. By the late 1920s every province had one.¹⁴⁶

Initially the fuel taxes were modest and seen as a supplement to licence fees. It did not take governments long to discover that a hidden user-tax like this was more lucrative and elastic than increases in licence or similar fees. As a hidden tax and as one that varied depending on the use of the vehicle, it also seems to have been more accepted. There was some outcry in Ontario when the first fuel tax was increased in 1925, but generally the motoring public seems to have been relatively willing to accept the burden.¹⁴⁷ By the end of the 1920s a series of tax increases and the growing number of cars on the road meant that fuel taxes surpassed licence fees (see Figure 8). Within a few years fuel taxes accounted for approximately three quarters of all user-impost revenue.¹⁴⁸

By the mid-1920s then, some important principles had been established for road policy. Most importantly, the new cost of roads had been met to a large degree by new taxes on road users. Over time the exact percentages of user-taxes devoted to highways would vary. In the 1920s, though figures

Figure 8
COMPARATIVE PROVINCIAL REVENUE FROM FUEL TAXES AND LICENSING FEES



Source: Statistics Canada (formerly Dominion Bureau of Statistics), *Canada Year Book*, 1925–1945.

are imprecise, it appears that road costs outstripped taxes. In contrast, during the Great Depression the cash-starved provincial governments spent less on roads than they took in. Then, after World War II, the percentages dropped again as massive road-building efforts were undertaken. In general, however, a rough correlation was maintained over the long term between user-tax revenue and expenditures. Occasionally this tie was required by legislation, but even where it was not the correlation existed as a matter of policy. Various commissions and investigations during the interwar years restated the principle that user-taxes on motorists should provide the basic tax support for the highway system.¹⁴⁹

Two other precedents were set in the 1920s. First, the roads were to be public property. This may seem obvious, but remember that railroads were not public property (except when bankruptcy left no choice) and there had been considerable use of toll roads in the 19th century. By the 1920s, however, toll roads had fallen out of favour and were gradually assumed by the appropriate government. In Ontario, for example, the last toll road was made free in 1926. It would be another generation before toll roads reappeared.

Second, though roads were clearly a provincial matter, the federal government had recognized the overwhelming importance of the system and become involved. The 1919 Act, originally intended to expire in 1924, was in fact extended to 1928. The Act set aside \$20 million of federal government money for cost sharing with the provinces. Each province would have access to \$80,000 automatically. The rest was to be divided up in proportion to population. To qualify for the money the provinces had to be meet standards of construction established by the federal government. These were not high by modern standards but do indicate the roadway system was being designed for cars rather than foot or horse traffic. The federal government also had a say over which roads were eligible, though any main road seems to have been accepted. In this age of criticism of anti-patronage sentiment, all contracts had to be let by tender. When these conditions were met the cost-sharing system would come into effect: the federal government would pay 40 percent; the provincial government, 60 percent.¹⁵⁰ The federal assistance does seem to have been important in the great construction effort of the 1920s. The provinces were provided with welcome financial assistance at a time when there was tremendous demand for new roads. The federal involvement also set a precedent. Thereafter provinces continually sought greater federal involvement in the highway system. The federal

government's involvement, however, has been intermittent: the Government oscillates between a desire to create a truly national highway system and a reluctance to get involved in a very expensive area outside its own jurisdiction.

The basis of the modern road system was established in the 1920s. In 1918 most of the system was still primitive. The only significant stretch of paved highway outside the cities was the concrete highway between Toronto and Hamilton. Only in parts of southern Ontario and in the area surrounding Montreal could one talk of a road system suitable for cars. Even then the roads were seasonal, closing down whenever the snow got too deep. In Nova Scotia, where the roads were better than in many provinces, the system was still sufficiently fragile that there was a provincial law on the books prohibiting motor traffic during the spring thaw.¹⁵¹ The side of the road the motorist should drive on varied from province to province, and British Columbia had only just asserted uniformity on that *within* the province.¹⁵²

The 1920s revolutionized Canadian highways in several ways. First, the technical obstacles to the development of a modern highway system were overcome. Roads designed for the horse and wagon were not suitable for the car. The speed and design of the car created new demands for gentle curves, limited grades and smooth surfaces. As well the wear and tear of cars were such that even a well-designed carriage road was unsuitable for car traffic. As the President of the Ontario Good Roads Association put it in 1914, "the automobile itself is quite a harmless instrument — a broad soft tire, but the rapidity at which it is driven upon our roads makes it a regular road smasher."¹⁵³ In the first 20 years of the century, failure to understand the implications of this fact led to numerous errors of planning. "Waterbound macadam," for example, reacted very differently under horses and cars. The slow, steel-rimmed wagon wheels compacted the macadam surface. The fast, rubber-tired car carved deep ruts into it.¹⁵⁴ In all areas, from drainage to gravel coverage to safe curvature, the car road had requirements considerably different from those of the older wagon road.

Second, until well into the 1920s only Ontario really had the demand or resources to create a full-fledged system of highway construction and maintenance. Other operations appear to have been relatively haphazard. In Alberta, for example, only one quarter of the maintenance appropriation was spent in 1922. Partly as a result of this, the legislature slashed the budget for 1923, leaving Public Works without the funds to do any but "the most urgent work."¹⁵⁵ That same year the entire maintenance cost for

the main Edmonton–Calgary highway was only \$6,961. The same general records of small expenditures, ad hoc arrangements and lack of long-term policy appear as well in the records of other provinces, or at least such other provinces as bothered to print annual reports on highway activities at all.

This changed dramatically through the decade. A 1923 report by the Ontario Minister of Highways said, “A much greater expenditure is to-day necessary and justifiable, as compared with the days when roads carried horse-drawn traffic alone.”¹⁵⁶ Certainly roads received a great deal more attention than they ever had before. With aid from the Dominion government, provincial highway budgets multiplied many times through the decade and became one of the major expenditures. Highway departments expanded and became much more involved in long-term development strategies rather than merely keeping existing roads passable. In British Columbia, for example, work began on the tremendously expensive trans-provincial highway, which, when finished would create “a serviceable highway from the Coast to the Crowsnest Pass.”¹⁵⁷ In Alberta the *Canada Highway Act* encouraged the province to create a high-quality north-south highway. By the end of the decade maintenance expenditures on it were 25 times higher than they had been in 1923.¹⁵⁸ In Ontario the emphasis was on the development of the King’s Highway system, including a considerable amount of hard surfacing. New Brunswick was starting from somewhat farther behind than Ontario, but it too put considerable funds into developing a main trunk system and improving rural roads.

The final change came in the response to weather. Until the 1920s roads were at the mercy of nature. Snow-removal equipment was primitive and the expenditure small. Ontario, for example, spent less than \$5,000 a year on snow removal through the early 1920s. Quebec spent less. The result was that roads, as always in the past, were in effect seasonal operations outside of the cities. Improved surfaces and drainage lengthened the season, but a snowstorm still shut the system down. Cars disappeared and sleighs took over. The roads “returned” to the 19th century.

As cars, buses and commercial vehicles became a standard part of the national culture and economy, however, such seasonality inflicted an ever greater social and economic cost. By the later 1920s all provinces were making more of efforts to create an all-weather road system. Ontario bragged that it maintained some 1,200 miles (1,900 km) of roads all winter (though snow could and did shut them for days at a time). Other provinces had neither the

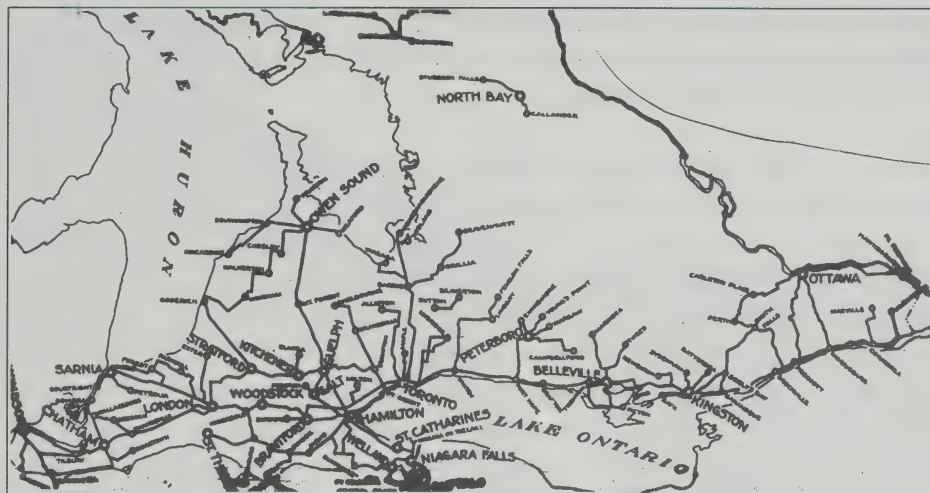
hard-surface mileage of Ontario nor the same money to spend on snow-removal equipment. Nevertheless, all provinces sought to keep at least the main highways open through the winter.¹⁵⁹ It would be another generation, however, before year-round car travel became the norm outside the cities of Canada.

What did this all mean by the end of the decade? What was the state of the Canadian highway system at the onset of the Great Depression (and it wouldn't change much for the next few years)? How, if at all, were railways affected by the rising popularity of the car? The most striking thing is the change that had taken place in a decade. There were still a good many bad roads, but the Canadian highway system had been redesigned and rebuilt for the car. Tracks had been turned into regularly passable roadways. Links between cities had been upgraded so that intercity automobile travel was routine, at least over short and medium distances. Nor was the car an exotic sight any more. Summer traffic on the Toronto–Hamilton road was upwards of 5,000 cars a day by 1926, and by the end of the decade the main road out of Vancouver was carrying more than 9,000 cars a day.¹⁶⁰

In contrast to the spread of the car, the use of the horse was declining. At the end of World War I roads everywhere had to carry both horse-drawn and motorized traffic. By the end of the decade the horse and carriage had become an increasingly rare sight along major highways, especially in the more urbanized east. Not only were there fewer of them, but as car numbers and rates of speed increased it took a degree of hardiness to bring a horse onto the main highways. Horses and carriages were now more or less confined to rural areas, and even there they tended to stay off the main roads.

Related to the motorization of the roadway was the development of commercial bus service. Through the 1920s bus service improved. In 1926 regular coach service began between Calgary and Banff. By the end of the decade the main cities of Alberta were served by scheduled bus lines.¹⁶¹ In Ontario, as Figure 9 indicates, an extensive system of bus routes had been developed, providing the first significant competitor to railways for the public passenger dollar. Railways, indeed, had now cast aside their earlier friendly attitude toward the car and were increasingly demanding that buses “pay their fair share.” Complex studies and adjustments of licence fees, gasoline taxes and related issues would become a major subject of discussion through the depression years as all modes of commercial carrier found their revenues in decline.¹⁶²

Figure 9
1927 Bus Routes In Ontario



Source: Ontario, *Annual Report of the Department of Public Highways*, 1927, p. 188.

For all the improvements that had been made in the road system, two facts that emphasize the limitations that still existed stand out. First, driving in the 1920s meant driving for the most part on gravel or unpaved roads. In Alberta, for example, the bulk of federal funds under the *Canada Highway Act* had gone to improve the main north-south route, Highway 2. By the end of the decade the province could boast of a “modern” gravel highway extending from Lethbridge to Edmonton. North of that it was still a rudimentary dirt road. A small section (70 miles [113 km]) was even paved in the early 1930s, but the Great Depression meant that most of it was gravel until after World War II.¹⁶³ The same conditions prevailed in Manitoba and Saskatchewan and through much of the Maritimes.

There were exceptions to this rule in the vicinity of the major urban centres. British Columbia’s roads on the lower mainland and southern Vancouver Island had hard surface by the later 1920s. So too did Quebec’s highways in the vicinity of Montreal. Ontario had by far the most extensive system, with some 1,800 miles (2,890 km) hard surfaced in one fashion or another as early as the mid-twenties.¹⁶⁴ The main highways westward and eastward from Toronto were paved, and it was possible to go all the way to Montreal

or Windsor on a paved highway. Even in Ontario, though, pavement was restricted to the major arteries. If one deviated from what is now known as the corridor, gravel became the rule. Only about 5 percent of Ontario highways were hard surfaced by the end of the decade. Moreover, even the paved highways were relatively narrow, with 18-foot (5 m) surfaces and gravel shoulders (see Figure 10).

At the end of the 1920s Canada's highway system was also limited by its regional nature. The *Canada Highway Act* had not tied grants to any sort of trans-provincial linkages, and provincial governments followed traffic demand rather than encouraging interprovincial links. At the end of the decade it was still impossible to drive across Canada. British Columbia was not connected to Alberta despite the efforts to create a trans-provincial highway. Ontario was not connected to itself: there was no road around the north of Lake Superior. Given the accessible American road system, the problem was probably more symbolic than real. More difficult and more costly was simply the patchiness of the Canadian road system. Except perhaps in the Windsor–Montreal corridor, long-distance travel by car was still difficult. Speed limits were still relatively low (they varied, but 30–35 mph (48–56 km/h) outside cities was typical). "Good roads" were still confined to the major transportation corridors and oriented on provincial rather than national needs. Whereas the railroad system had strong nationalist associations, the road system was, after the first 30 years of the car, regional in orientation and uneven in standards.

The car was sufficiently important by the 1920s, though, that the railways began to complain about "unfair competition." "Motor vehicle competition by private automobile, bus and motor truck," complained the CNR in 1931, "has made serious inroads into the railway's traffic. . . . The use of the motor vehicle beyond its proper economic sphere can have no other effect than to increase the total cost of transportation to the citizens of Canada."¹⁶⁵ That "proper sphere," apparently, did not include competition with railways. Concern focussed on two areas in the interwar years. The first, not relevant to this paper, was short-haul freight: commercial trucking firms entered into competition with railways quite early and quite effectively. The second is relevant, however: the railways saw the car and bus as just as much a threat to passengers as to freight.

Figure 10
No. 2 PROVINCIAL HIGHWAY, 5 MILES EAST OF BOWMANVILLE



Source: Ontario Archives, Negative No. A0913 (reproduced in Ontario, *Annual Report of the Department of Public Highways, 1926*, Frontispiece.)

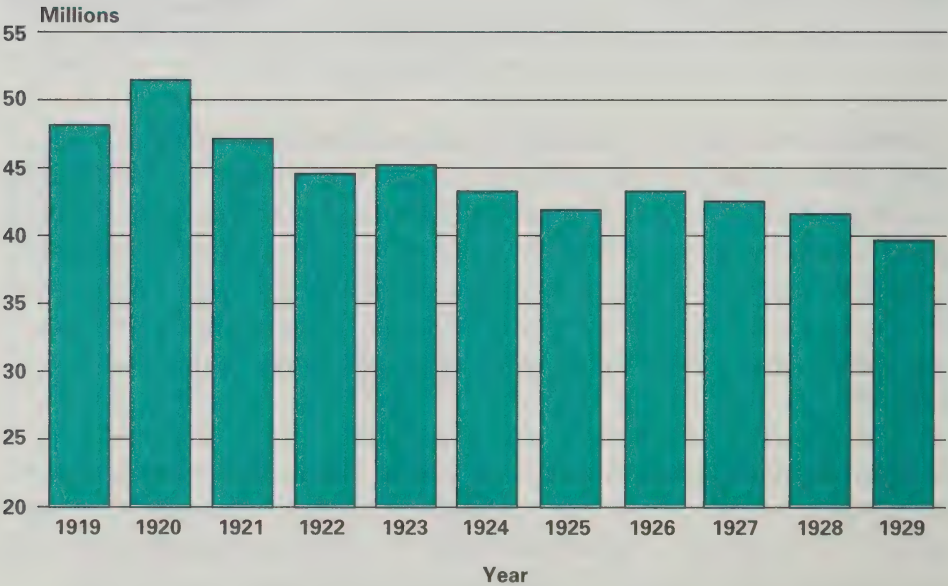
The main concern was the attractiveness of the car to the individual or family going a relatively short distance. A trip from, say, Hamilton to Toronto or Brandon to Winnipeg by horse and wagon was not competitive with the speed, comfort and even cost of the train. The same could not be said when it came to a comparison between the train and the car. Even more threatened was the very short haul. Why travel according to someone else's schedule over a distance of 10 or 15 miles when you could simply start the car and drive off?

An historian of the Canadian national system confirmed the concerns of railways, noting that the "motor car had taken a heavy toll of railway passenger traffic."¹⁶⁶ The statistical material available also shows some support for the railways' fears. Though the 1920s were generally good ones for the railways, several indicators imply either a static or somewhat declining

passenger market. As Figure 11 indicates, for example, there was a slow decline in the number of railway passengers through the decade despite improving economic conditions after 1921 and a growing population. Figure 12 also gives at least indirect evidence that there was a correlation between the rise in cars and the erosion of passenger numbers on the railways.

The competition from cars and buses at any time between the wars should not be overstated, however. For one thing there was simply not enough of them. As mentioned, even by 1929 there was only one car for every nine Canadians. For the great majority of Canadians, public transportation was still the standard means of travel. Even for those who had cars, the technology and infrastructure of motoring were not yet adequate for regular travel over anything but the shortest distances. Cars had come a long way in the decade, but they were still designed for relatively low speeds and short distances. Suspension systems were rudimentary. Most cars were not heated. For that reason as well as the primitive nature of snow clearance, much motoring was a seasonal phenomenon.

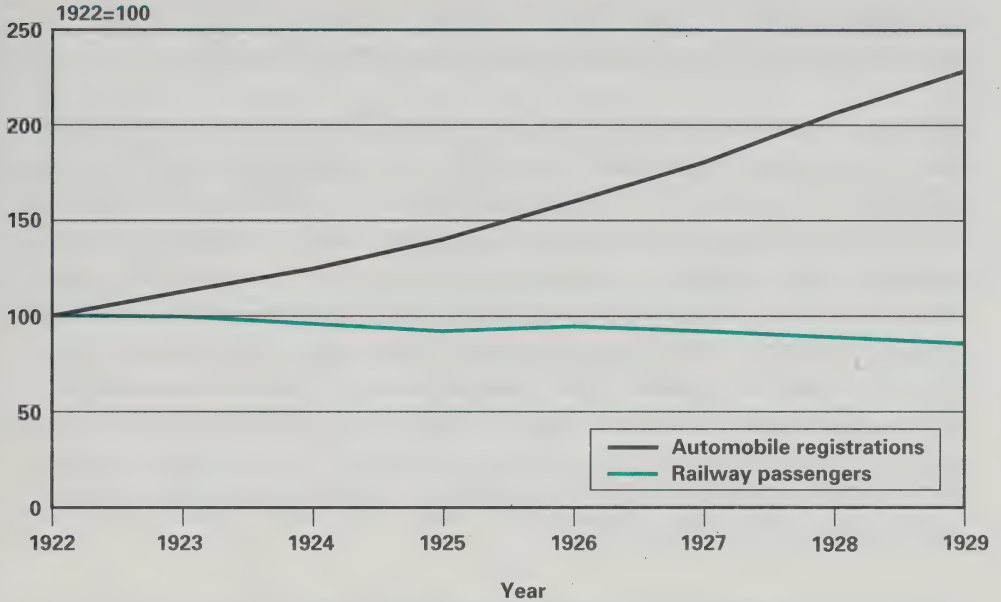
Figure 11
NUMBER OF RAILWAY PASSENGERS CARRIED, 1918-1929



Source: Statistics Canada (formerly Dominion Bureau of Statistics), *Canada Year Book*, 1930, p. 626.

Figure 12

RAILWAY PASSENGERS AND AUTOMOBILE REGISTRATIONS, 1922-1929



Source: Statistics Canada (formerly Dominion Bureau of Statistics), *Canada Year Book*, 1930.

Between the wars the most serious challenge to the railway came in a specific area, short-haul commuter traffic. It was the electric and street railway, designed for brief trips and constructed in and around those urban areas where roadways were usually best, that faced the brunt of the competition. Electric railway mileage peaked in the late 1920s, and thereafter abandonment rather than construction was the trend. By the 1950s many cities had totally abandoned their electric rail systems. Others had only a vestige of this once crucial urban and commuter link.¹⁶⁷ Across greater distances, however, the car was but a minor threat to the dominance of the railways in passenger travel. The commercial airplane was such an insignificant factor before World War II that all discussion of it is best deferred until after 1945.

THE GREAT DEPRESSION

In 1929 the economy turned downward as world commodity prices collapsed. Canada, as a major exporter of such commodities, was quickly affected. The Prairies had a bad year in 1929, and by 1930 the economy was in trouble across the country. For the next three years the Canadian and world economies

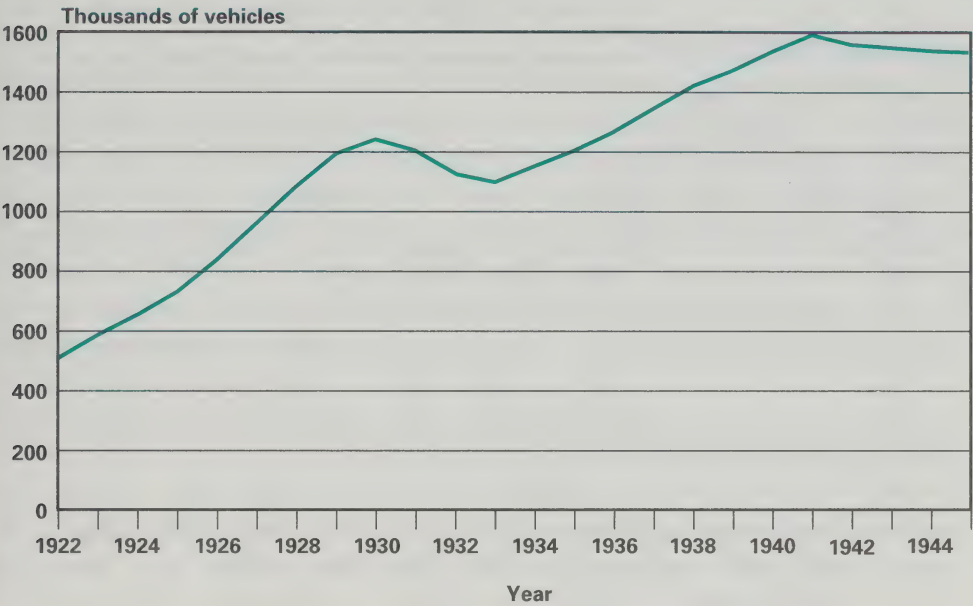
would continue to shrink until, at the bottom, world trade was at only one third of the level of 1929. Then, for the rest of the 1930s, there was only slow and intermittent improvement. Not until the World War II was the Canadian economy prosperous once again. In the intervening years all aspects of Canadian passenger transportation were affected.

The Great Depression rapidly brought the highway boom of the 1920s to a halt. This is hardly surprising. The depression shrank the financial resources available to governments. In 1930 provincial revenues had reached an all-time high of \$188 million. By 1933, the low point of the depression, they had dropped to \$152 million, or by almost 20 percent.¹⁶⁸ At the same time provinces were being pulled by bankrupt cities into more and more involvement in relief programs. As a result deficits rose and so too did the pressure to cut costs wherever possible. Road programs were a natural target. For one thing, government cutbacks are often carried out at the expense of the infrastructure. Capital projects can readily be delayed. So too can maintenance, which can be minimized with the recognition that a delay of a year or two will unlikely do much harm.

Another factor reinforced the tendency to delay capital construction. With the depression, automobile ownership ceased to increase at anything like the rate it had before. Indeed, for a period between 1931 and 1934 absolute automobile ownership decreased in Canada (see Figure 13). This would happen again, though at a lesser rate, between 1941 and 1945 as wartime restrictions curtailed production. Even when registrations did turn upward in the later 1930s the rate of growth was much slower than in the 1920s (see Figure 14).

At a time when governments were starved for revenue this decrease in the rate of growth of cars for some 15 years (and absolute drop in ownership for part of the time) provided a welcome respite. Capital expenditures could be slashed not just on the basis of fiscal necessity but because of decrease in demand. Maintenance was in theory less affected by the halt in the growth in the number of cars, but that too was cut back severely. Revenue decreases, said the 1932 *Alberta Annual Report of the Department of Public Works*, "made it imperative to curtail highway expenditure to such an extent that the utmost economy and watchfulness became necessary." The *Ontario Highways Report* for 1933 referred to a decrease in activity both in maintenance and construction because of "the depressed state of the financial economy."¹⁶⁹ It was the same everywhere as provinces finished off old projects and postponed or cancelled new ones.

Figure 13
REGISTERED MOTOR VEHICLES

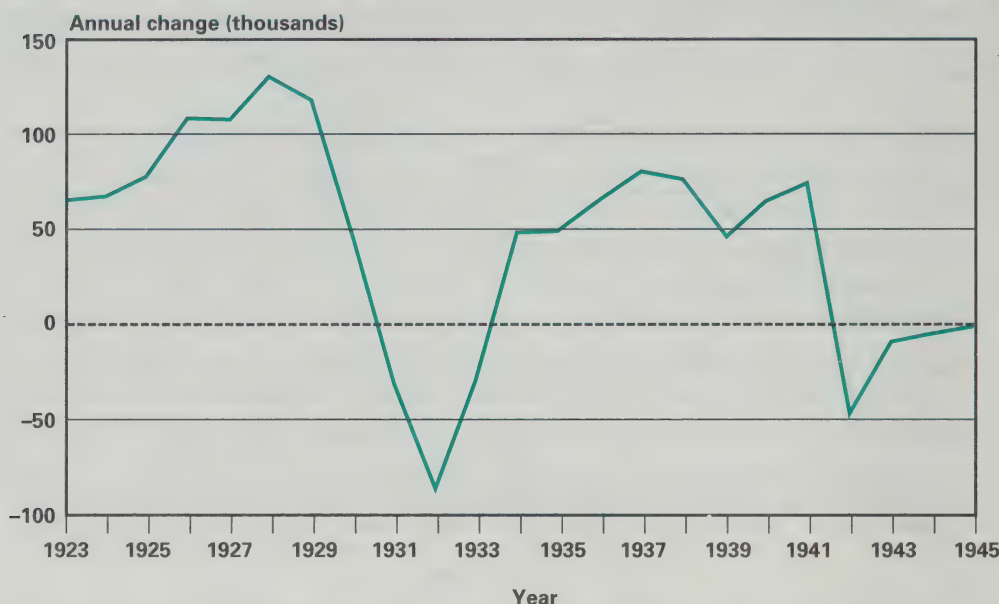


Source: Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series T147-194.

The overall decrease in provincial expenditure on roads was dramatic, declining by 60 percent between 1930 and 1933. Thereafter expenditures increased, but it was 1937 before highway expenditures recovered to previous levels. In the meantime user taxes on cars actually increased: cash-starved governments raised the vehicle registration fees and fuel taxes repeatedly through the decade. At the end of the 1920s user fees had recovered only about 50–60 percent of the total provincial highway expenditure in Canada. In five of the 10 years of the Great Depression user revenue matched or exceeded total provincial highway expenditure. User taxes were now subsidizing general revenue rather than vice versa. This pattern would become even more dramatic in World War II as road taxes outstripped expenditures during every year of the war, sometimes by as much as 50 percent.

Hard as the roads were hit, the railways were hit harder. The depression affected them almost immediately. Revenues dropped in 1929 and again in 1930 as wheat shipments declined from the Prairies. Then, as the depression hit home, people travelled less on business or pleasure, and passenger

Figure 14
MOTOR VEHICLES REGISTERED



Source: Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series T147-194.

traffic plummeted. The joint operating revenue of the two systems decreased from \$469 million to a low point of \$262 million, a decline of 45 percent. In response, both railway systems scrambled to avoid collapse. They reduced their operations, drastically slashing the number of trains, maintenance levels and the work force. By the low point, activities had been cut to the point that it was as if one of the railway systems had disappeared.

The two railways were affected differently. CPR was better able to weather the storm, though it had to suspend dividend payments for the first time since the completion of the transcontinental line and even had to turn to the government for some short-term debt guarantees.¹⁷⁰ Things were worse for CNR. The fixed debt CNR had inherited from the Grand Trunk and Northern railway systems meant that even in good years it was difficult for it to meet expenses and debt charges. To make matters worse, CNR was not really a coherent network of railways. It was a patchwork, put together from all those overextended, non-viable lines the government had assumed over the years. Finally, CNR had been as aggressive as CPR in adding to Prairie

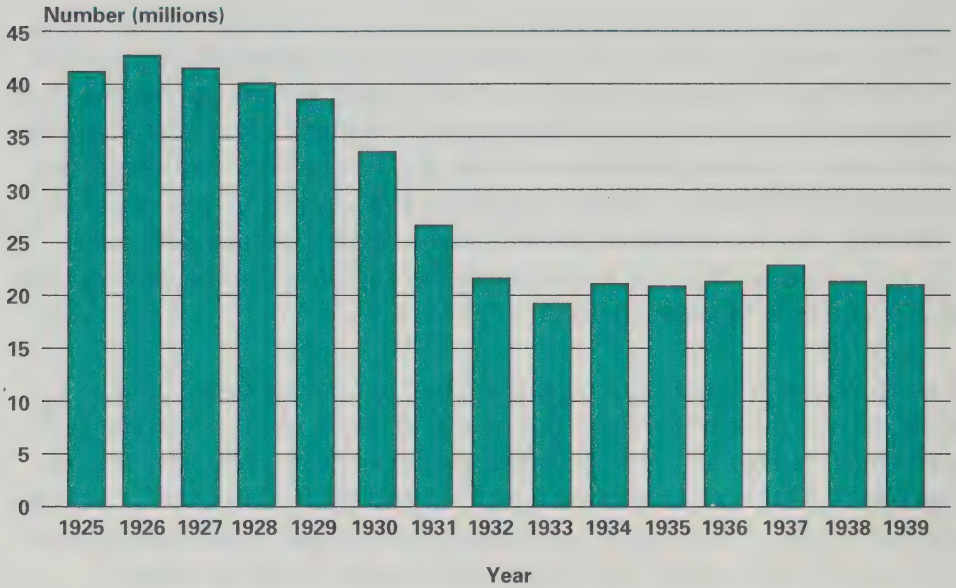
branch lines through the 1920s, which of course added to the debt as well.¹⁷¹ At the beginning of the Great Depression CNR's debt was massive, equivalent to one quarter of the national debt.

CNR's financial position made it especially vulnerable to the depression. In 1929 CNR had an operating revenue of nearly \$260 million, against operating expenses of just under \$215 million; the profit therefore was approximately \$45 million, excluding payments on debt. By 1933 revenue had declined to less than \$150 million; operating expenses, to \$143 million. The ongoing operation was just a little more than half the size it had been before the Great Depression.¹⁷² In the same period, the number of employees declined from more than 111,000 to 70,000.

Passenger systems played an important part in both the decline in revenue and the reduced services. During the 1920s, as we have already seen, passenger-related services like hotels had been major expenses for both systems. A 1931 royal commission used the benefit of hindsight to condemn the two railways for creating "a standard of passenger travel quite beyond the needs of the country" and for excessive expenditures on hotels.¹⁷³ Then, after all that expenditure, passenger travel collapsed. As Figure 15 indicates, the decline in ridership continued until 1933, by which time the railways were carrying half the passengers they had in 1928. Thereafter the decline halted, and there was even some improvement. Even after a decade, however, passenger traffic had improved only marginally from its low point and was far below the level of the 1920s.

Passenger revenue declined even more sharply than passenger numbers. As Table 2 shows, between 1929 and 1930 CNR experienced a 14 percent drop in passenger revenue, with another 29 percent drop in 1931. People simply could not afford to travel any more. Sir Henry Thornton, President of the company, had always been interested in passenger service. Nevertheless, the company was desperate and immediately moved to slash expenses. New car orders were delayed. Some of the more expensive luxury services (for example, parlour cars, sleeping cars) were cancelled on various runs, and services themselves were slashed. The number of passenger-miles decreased by 480,000 in 1930, more than 3 million in 1931 and close to 5 million in 1932. The CNR *Annual Report* commented that this was "equivalent to wiping out practically one-third of the passenger train service operated in 1929."¹⁷⁴

Figure 15
THE DEPRESSION AND THE DECLINE IN RAILWAY PASSENGERS



Source: Derived from Canadian National, *Annual Report*, for the years 1926–1939.

Table 2
CANADIAN NATIONAL RAILWAY PASSENGER REVENUE, 1929–1940

Year	Passenger revenue (\$000)	Passenger revenue as % of operating revenue
1929	32,013	12.32
1930	27,537	12.42
1931	19,657	11.09
1932	17,259	10.71
1933	15,032	10.12
1934	17,554	10.65
1935	17,863	10.31
1936	17,022	9.12
1937	18,945	9.55
1938	18,097	9.93
1939	17,817	8.74
1940	21,702	8.77

Source: Derived from Canadian National, *Annual Report*, for the years 1930–1940.

Some writers have looked back on the Great Depression as a fundamental turning point in the history of passenger transportation on railways. Until then, this argument goes, the railways had seen passenger transportation (and ancillary services) as an important part of their revenue and as a matter of prestige as well. Then, with the depression, the whole emphasis on passenger service shifted from growth to retrenchment and from prestige to liability. Rail passenger service never recovered thereafter.

There is some evidence for this. There is no doubt that the Great Depression brought a tremendous reduction in service. A good many marginal routes were cancelled; services on others were cut back. The public, which had become accustomed to ever-improving regularity and proximity of passenger services, now experienced the reverse. Even as the Great Depression lifted and some services were restored, rail transportation was not as it had been before. Many of the routes were not revived, and on more major runs there were fewer trains scheduled than before. Even well into the boom times of the World War II, for example, passenger-miles were well below what they had been in the 1920s. As we have seen, both railways trimmed operating expenses by reducing capital expenditures and deferring indefinitely much-needed maintenance on tracks and equipment. Trains became a little more infrequent, services a little more difficult to get at and the level of comfort a little lower. As we will see, the war provided the railway with certain advantages, and a sharp increase in ridership resulted. The conditions during the war were abnormal, however, and once it ended the railways found themselves facing more competition than ever. The decade of the Great Depression and the reduction in equipment and service did much to undermine the railways' ability to meet that competition.

Another reason to think of the 1930s as a turning point comes when one looks at the relative importance of passenger traffic on the system. For it was in the 1930s that passenger traffic on both major lines began to slip in importance. As Table 2 shows, CNR's passenger traffic accounted for 12.32% of revenue in 1929 and only 8.74% a decade later. The same relative decline occurred on the CPR: 6.7% in 1929 and only 4.8% a decade later.¹⁷⁵ When these figures are put together with the lesser declines in the relative importance of passenger revenue of the 1920s, the cumulative trend is quite significant. Passenger services were becoming a peripheral part of the railway operations.

Finally, the 1930s were a turning point because of the changes in the relationship between the railways and the Government. Not surprisingly, given the importance of the railways to the national economy and the tremendous burden that CNR placed upon the Dominion treasury,¹⁷⁶ there was a great deal of controversy surrounding the beleaguered railways in this decade. There were parliamentary hearings, the already-mentioned royal commission and some significant alterations in Government regulatory practices. Sir Henry Thornton, the pro-passenger head of CNR, was removed and CNR was placed first under a board of commissioners and then returned to its traditional form of governance. In 1936 the Department of Transport was formed, replacing the old Department of Railways and Canals.

Perhaps the most important change in passenger transportation arose from the growing concern of the Government. It asserted with increasing force its right to intervene in railway business decisions with an eye to the public good. Of course, as we saw earlier, this precedent was well established under the common-carrier rule. Since 1903 the Board of Railway Commissioners had sought to protect passengers' basic health and safety as well as regulate the passenger fare system. Still, at the beginning of the 1930s, railways made the decisions about what trains to run and whether they should be run at all. The Minister of Railways and Canals summed it up in 1933 when queried about cutbacks in passenger service in northern Ontario. "Such matters as running trains are not discussed in any shape or form with the government. That is entirely a question of railway management."¹⁷⁷

Such attitudes were feasible as long as the overall railway passenger system was healthy and, for the most part, expanding. The Great Depression destroyed any such assumption. Passenger service was decidedly no longer a growth industry. As with railway service generally, the tendency was now toward a reduction rather than improvement in service. The government increasingly feared, however, that in the panic of railways to reduce expenses the welfare of the public would be ignored. In such circumstances the Government felt bound to step in. Thus in 1933 it passed a new statute requiring railways to get the approval of the Board of Railway Commissioners before abandoning lines.¹⁷⁸ This stopped short of Government supervision of specific passenger services, but it was a significant change in the previous assumption that such matters were best left to the companies involved. Over the next few years both CNR and CPR regularly applied for permission to abandon lines and, on average, won the right to do so in about half the

cases. Thus the Great Depression set precedents for dealing with a new reality — declining service. The Government extended the precedent of the common carrier by asserting its right to control branch-line abandonment. The railways, for their part, found themselves in the uncomfortable position of being forced to operate unprofitable lines. Though these precedents applied to abandonment, the practices could and would be extended to the provision of specific services within a few years.

The railways did provide a cheap service for a population that could afford little more. CNR boasted in 1934 that standard fares amounted to only 2.25 cents a mile (1.4 cents a kilometre).¹⁷⁹ Along the way, however, the sharp drop in revenues forced CNR to reduce the choice of trains and the level of service. By the end of the 1930s the train was not as accessible or convenient a mode of travel anymore. Also, despite some improvements such as a limited introduction of air-conditioned cars after 1935, passenger equipment was allowed to run down in the 1930s. The war only exacerbated the problem so that by 1945 equipment was increasingly unreliable and shabby. Massive overhauls and purchases would be needed to bring it back to the standards of pre-depression days. Finally, all of this was happening as the relative importance of passenger service was declining. Passenger service, once the most prestigious part of a railway's business, was not yet seen as an undesired burden, but it was increasingly an operation of peripheral importance.

WORLD WAR II

World War II saved the railways from imminent collapse. The Great Depression came to a rapid end as the nation produced the goods for a modern industrial war machine. On the Prairies wheat harvests recovered along with world wheat prices, and grain shipments on the railways recovered in parallel, first to pre-depression levels and then to new records. The operating revenues of both railways shot up even more dramatically than they had fallen at the beginning of the 1930s, increasing by 67 percent between 1939 and 1941.

Initially, in a trend that continued from the late 1930s, most of this improvement was in freight services. In the case of CNR, for example, 1939 saw freight revenues increase by 14.6 percent, compared with a 1.5 percent decrease for passenger revenue.¹⁸⁰ It was not long, however, before passenger travel

also boomed. The number of passengers carried by the railways tripled between 1939 and 1944 (the peak year). Revenue expanded accordingly, reaching all-time highs in both current and constant dollars.¹⁸¹ By the end of the war some 60 million passengers rode the trains — five trips for every Canadian citizen.

Much of the boom in passenger travel during World War II was a short-term anomaly, for the recovery of passenger service, spectacular though it was, did not represent a return to a more competitive and hence more profitable position for the long term. With the one exception of 1944, passenger traffic did not recover to a point where it could match the ridership of the 1920s. Railways did as well as they did in the war years for very specific reasons. Car transportation was severely restricted by two factors. Canadians had not replaced their cars very quickly during the depression years. For a time registrations had actually decreased, and even toward the end of the 1930s people tended to hang onto the existing vehicle, make one more round of repairs and hope the thing would run another winter rather than go into debt or squander precious savings to buy a new one. By the beginning of World War II, therefore, Canadian cars were rather old overall. Undoubtedly there would have been massive replacement of them in the prosperity of the war years except that production of new cars effectively ceased by 1940 or 1941. No doubt there was tremendous demand for cars, but they were simply not available. Car registrations actually declined during the World War II, as did the number of miles driven.¹⁸²

There were additional restrictions. Gasoline rationing began in 1942, thus limiting the amount of driving owners of cars could do. Even commercial activities were restricted. Intercity bus trips were restricted to no more than 50 miles (80 km).¹⁸³ Taxi travel between cities was also effectively prohibited. Airline travel was still in its infancy, and during the war regular flights were practically impossible to get. Thus, for the duration of the war, there was no real commercial competition for the railways over longer distances: this was a time when wartime business, troops coming home on leave, and relatives visiting troops on station created more than enough traffic to make up for the absence of immigrants or tourists. The formal railway rates for passenger services were not increased during the war because the rates, as with much else in Canada, were frozen under Wartime Prices and Trade legislation. Nevertheless, railways were effectively able to increase revenue by dropping the deeply discounted excursion fares that had been used to lure customers during the Great Depression.

The wartime boom in passenger travel, therefore, did not mark a return to the golden pre-depression years. Yet it was important not just because it rescued railway passenger services from insolvency but because it introduced a new generation to train travel. The train took the soldiers off to war, and the train brought them back home. The train took the young bride off to Halifax or Esquimalt or elsewhere to meet her husband on shore leave from the Navy. The train moved the war materials and responded efficiently to the crisis of these years. Once again, the train's role as a part of the sense of national identity was reinforced.

5. THE COMPETITIVE ERA AND THE RISE OF THE PASSENGER POLICY ISSUE: 1945-1967

... we find that there is little social justification and less economic, for the permanent provision of railway passenger services as we know them today. The public, by and large, has already indicated its preference for other modes of travel. ... (MacPherson Commission, 1961)¹⁸⁴

Until sometime around World War II, passenger transportation in Canada was dominated by the railway companies. The car had made inroads, but these had been limited by the Great Depression and by wartime restrictions on production. In contrast, the post-war era brought the end of dominance by railways as the car achieved hegemony over shorter distances and the airplane over longer ones. As this happened the railway passenger travel networks became extremely costly to maintain. From the later forties onward the companies sought to reduce or eliminate specific services and increasingly looked for internal forms of cross subsidization and then direct Government support. Politicians and the public, who until then had been focussed on freight rates, increasingly turned their attention to the passenger issue.

Many services have disappeared with minimal fuss when outcompeted by newer forms. Lake- and ocean-going passenger ships disappeared, as did most streetcars. Aficionados lamented their passing, and a few editorials complained about the disadvantages of the loss of service. Generally, however, neither Government nor the public expected such transportation systems to remain beyond their competitive life. Passenger railways were and are a different matter, however. The railway was so vital in the recent past

of the nation and its mythology so entwined with national history that abandonment could not be considered just another step in the evolution of technology. Further, the well-entrenched doctrine of the common carrier meant that there were numerous precedents for treating the railways as a public commodity. There were also a rising number of precedents for subsidization of railway services in the name of some national or regional goal. Currents of nationalism from the mid-1950s onward complicated the issue. The abandonment of the railway seemed tantamount to weakening Canada. Regional balances came into question as outlying communities and smaller urban centres complained that the destruction of rail service was a betrayal of their place in the nation. All of these debates occurred against ever-mounting passenger rail losses and a more competitive transportation network.

In rough terms the post-war era can be divided into two parts. The first part, from 1945 to the mid 1960s, was one in which passenger transportation and passenger transportation policy was in a state of uncertainty. The government, faced with political pressure on issues like freight rates and line abandonments, did not really look at passenger transportation as a distinct problem. In the meantime travel by car and air was expanding rapidly, changing the economics of passenger rail more quickly than the railways could adjust their policies or practices. The railways were still interested in providing passenger service, but both they and Government seemingly underestimated the changes that would be needed to cope with the new conditions of the post-war situation. The report of the 1961 *Royal Commission on Transportation* (MacPherson Commission) and the passage of the *National Transportation Act* (1967) mark a rather drawn-out transition to the second post-war era. By this time the Government recognized that the previously powerful passenger rail system was less able to compete with other modes. The Government sought to deal with the situation by allowing a combination of abandonment and subsidy, depending on social and economic circumstances. Thereafter passenger rail became a subsidized system in which various policy steps attempted to balance public demands for ongoing rail transport against governmental desires to reduce subsidies and rationalize the system.

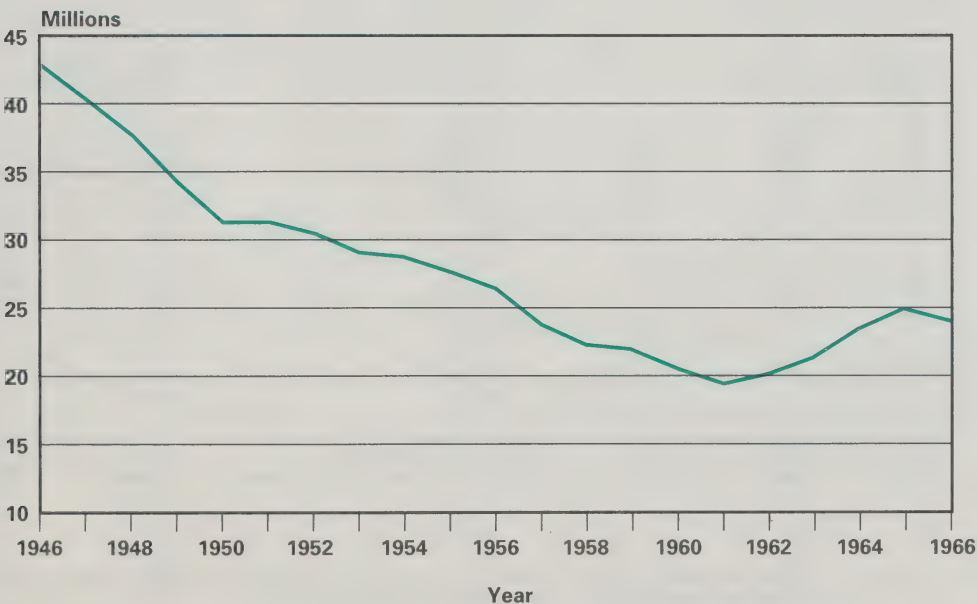
Two things link these distinct periods. The Government increased its activity and investigation in the area of transportation in the post-war period. There were two royal commissions on railway problems in just over a decade. In between there was a third (the Gordon Commission),¹⁸⁵ which dealt with the issue as a part of Canada's economy. Freight and passenger rates, which

had been more or less stable for a generation, were increased several times, then reduced, and then subsidized as Government sought a solution for the changing situation. Subsidies multiplied in size and diversity.¹⁸⁶ Through all the changes both the Government and the public wrestled with two difficulties. The first problem, as a post-war transportation economist put it, was whether "transportation services to be regarded as business institutions like department stores, factories, or farms, or are they to be looked upon as almost eleemosynary agencies wherein the cost-revenue relationship is subordinate to the welfare of the public?"¹⁸⁷ The second problem was to recognize that a system of regulation that had grown up because of the power and effective monopoly of the railways now had to be readjusted because of their weakness in the face of cars, buses, and airplanes.

THE RISE OF COMPETITION IN THE POST-WAR ERA

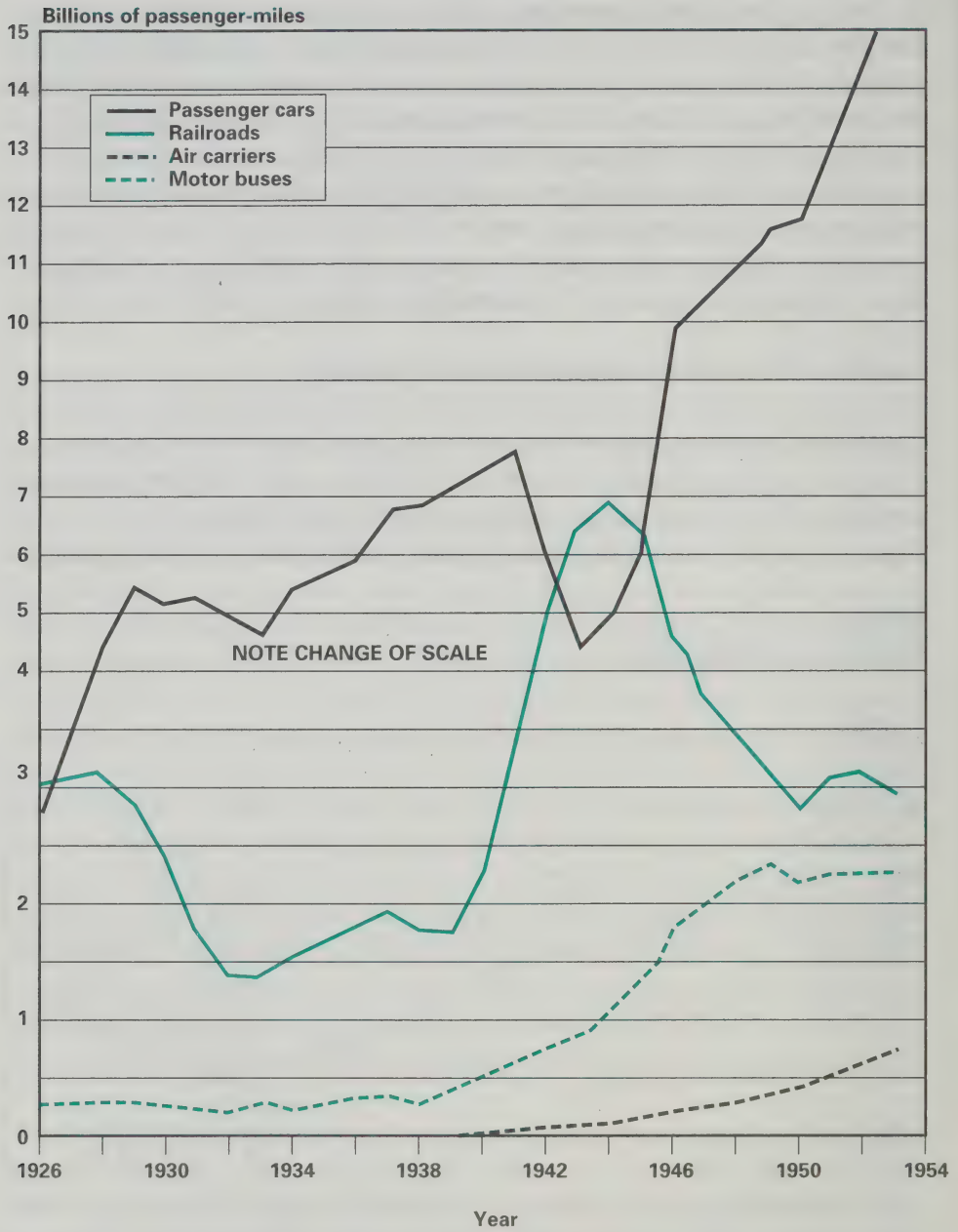
The decline of railway control of land travel after World War II is readily demonstrated. Every measure indicates the same story. After the records set during the war, passenger travel by railway dropped sharply. As Figures 16 and 17 show, the railways quickly found that the ridership levels

Figure 16
NUMBER OF RAILWAY PASSENGERS



Source: Statistics Canada (formerly Dominion Bureau of Statistics), *Canada Year Book*.

Figure 17
ANNUAL VEHICLE-MILES OF TRAVEL IN CANADA, 1926-1953



Source: Gordon Donald Campbell, "An Analysis of Highway Finance and Road User Imposts in Canada," PhD thesis, Purdue University, 1956, p. 62.

of the war years were artificially high. Conversely, car miles increased dramatically once wartime restrictions disappeared. A steady downward trend in both the absolute and relative position of railways began that continued until, by the end of the 1950s, ridership levels were returning to those of the depression era.

Revenue followed numbers. Railways could not increase fares to make up for declining ridership because that would only accelerate the desertion to other modes of transportation. Instead, to keep customers they had to cut fares. Excursion packages, discounts and group rates were used to woo passengers. Some of these helped, but the general trend was irreversible. This is especially striking because revenue was falling, even though Canada's population was rapidly growing as a result of the famous baby boom and high post-war immigration. As Table 3 indicates, the population was using railways much less frequently on a per capita basis. Travelling by train was a much less common experience by the 1960s than it had been even 15 years earlier. The long-term trend was even more dramatic. Average Canadians of the early 1960s took only one fifth as many trips as their grandparents had in the 1920s.

Table 3
ANNUAL RAIL PASSENGER REVENUES AND RIDERSHIP, 1946-1966

Year	Passenger revenue (millions of 1971 dollars)	Passenger trips per Canadian
1946	222.22	3.53
1947	178.46	3.26
1948	148.31	2.99
1949	146.72	2.60
1950	131.66	2.27
1951	134.85	2.21
1952	134.17	2.09
1953	128.36	1.93
1954	121.81	1.86
1955	122.96	1.73
1956	124.53	1.62
1957	122.91	1.38
1958	106.47	1.25
1959	100.27	1.20
1960	93.14	1.09
1961	81.60	1.03
1962	79.84	1.04
1963	77.07	1.09
1964	81.04	1.19
1965	81.61	1.25
1966	75.21	1.16

Source: Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series T44, T62.

In the 1930s the railways had suffered along with the rest of the economy. This time revenues were declining despite national prosperity. With the exception of only a couple of short-term recessions, Canada underwent steady economic expansion from the war's end until the later 1960s. In the case of railway travel, however, the economic boom only exacerbated the difficulties of passenger transportation, for it enabled the more rapid introduction of alternative forms of technology, the car and the airplane.

CARS AND ROADS IN THE POST-WAR ERA

At the end of the war the desire for the car presented a classic case of pent-up demand. The war years had brought full employment, and the nation had quickly climbed out of the depression. Under normal circumstances the aged fleet of cars possessed by Canadians would have been turned in over a period of years. New buyers would have entered the market as economic circumstances warranted. As we have seen, however, car production was halted during the war. It was thus the railway that reaped the benefits of wartime travel. The demand for the car continued to build, however, and in the post-war period the car was no longer a luxury but a necessity.

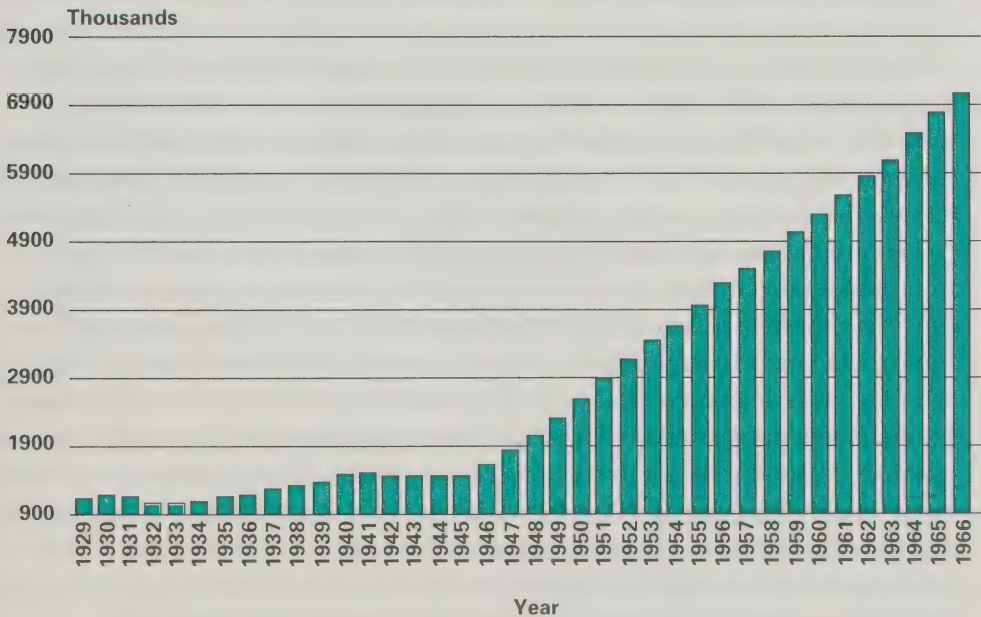
The only thing that was required to release this pent-up demand was prosperity. That came. The period from 1945 through the later 1960s was one of the most prosperous in Canadian history. The average Canadian's standard of living would double in these years, and one of the things that Canadians saw as an absolutely necessary part of that rising standard was a car. After more than 15 years of practically no growth Canadian car registrations exploded following the war (see Figure 18). In 1946, despite limited production of cars, registrations increased by 74,000; in 1947, by 136,000. By 1952 registrations were double those of 1945. It was estimated that more than 50 percent of Canadian families owned a car by then. "In other words," said one contemporary observer, "it required only seven . . . years to duplicate the motor vehicle production and ownership achievements of the previous 45 years."¹⁸⁸ That was just the beginning. A decade later car registrations had doubled again. In 1945 one in eight Canadians had owned a car. By 1954 it was one in four and by 1965 one in three.

Patterns of ownership also changed. First, class ceased to be a serious barrier to car ownership. By the 1960s all but the very poor owned a car, and other factors, such as age or disability, were more likely to prevent car

ownership than income. Second, regional variations in per capita ownership decreased. From the beginning, Ontario had had more cars per capita than any other province. In 1945 that was still the case: Ontario had more than twice as many cars per capita as, for example, Quebec. In contrast, by the mid-sixties the differences between provinces were relatively small.¹⁸⁹ The car had become a mass phenomenon.

Critics have railed against the car of the 1950s. They have pointed out, with some reason, that it was overpowered, ridiculously unsafe and a polluter. Such criticisms did not reflect the popular mood. North Americans felt a tremendous attachment to their cars in the post-war decades, for however far it fell short of the ideal, the 1950's car was a great improvement over what had gone before. It was more powerful, smoother riding, heated and, after the mid-1950s, often equipped with a radio. The average worker of 1955 could afford a car immeasurably superior to that of his parents. Moreover, for the suburban family the car was essential. No other alternative was competitive in time or money for the thinly populated suburbs. In other words, the love affair with the car starts from a practical base.

Figure 18
AUTOMOBILE REGISTRATIONS



Source: Derived from Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series T147-194.

The 1950's car wasn't just a practical appliance. The decades immediately after the war were dominated by the belief that what was modern was good. The old dusty ways of the Victorian era or, more recently, the staleness of the depression were to be swept aside by new technologies and new designs. The old represented poverty at home and appeasement and lack of confidence abroad. The new ways and new technologies symbolized the massed technological inventiveness of Western society — the same inventiveness that had won the war.¹⁹⁰ This was a technological age, and the car was the item of technology closest to the average citizen.

Modernist enthusiasm was focussed as much on design as on technology. There was a belief that post-war democracies were building a new society from the ruins of war and depression. The ranch-style bungalows, the modernist buildings that dotted the cities and the grand physical projects like the St. Lawrence Seaway and Trans-Canada Pipeline reflected the faith that progress came through modern techniques applied to the physical environment. Design reflected what has aptly been termed the "populuxe syndrome" — imitations of luxury on cheaply designed mass-production items.¹⁹¹ There was also an element of insecurity beneath all the rhetoric about progress and development. The ownership of family assets meant a great deal to a generation of Canadians brought up in the substandard housing of the Great Depression or in the overcrowded housing market of war.

The appeal of the car was that it so neatly combined these two deeply felt characteristics of the age — faith in technology and insecurity. On the one hand, the car, with hydro-glide transmissions, swept-fin styling and a great deal of chrome, was the primal example of the populuxe age run rampant with such suggestive names as Strato-Chief, Rocket '88, Bel Air and Biscayne, it was deliberately associated with both technology and luxury. On the other hand, the car reflected the desire of the post-war generation to raise their families with a degree of comfort and security unknown when they themselves were children. "Give Your Family Big Car Dependability," ran one Chevrolet ad, with a picture of parents and two children cruising along the inevitable open road.¹⁹² Ford blatantly sought to capture the hopes and anxieties of the age when it wrote its own version of the history of the car in a one-page ad entitled "Escape to the Greenbelt." It was a story of inner-city decay and violence contrasted with suburban peace and tranquillity. "The whole population is moving from the stone and steel of the city toward the fresh air, the light, the trees and living space of the suburbs."¹⁹³

As the above ads indicate, car advertising was selling a way of life. Practically without exception the car and its proud owner were depicted in country settings. Usually there was a married couple, occasionally with others present as well. The settings were always pleasant, often in the fall (when the new models came out) and often associated with such activities as picnics and sports. Typical was a later General Motors ad depicting a man fishing in an idyllic brook in a wooded glen. On the bank of the stream is his new Chevrolet. "When George Martin goes fishing," reads the copy, "he gets away from his worries. He forgets about missiles [this was 1961], taxes and the other problems that are a part of daily life."¹⁹⁴

The road, the corollary to the car, appealed to many of the same values. It too was the product of technology and sought to transform the quality of life through physical development of the infrastructure. When the Gardner Expressway was first announced by Metro Toronto Council, the *Telegram* ran a one-word headline: "Whoosh."¹⁹⁵ The speed and freedom of the car meant nothing without the road designed to handle it.

On a more practical level, highway traffic had now become so important that good roads were an important aspect of economic prosperity. Indeed, the rhetoric about the importance of a good road infrastructure is reminiscent of earlier debates about railways. A 1956 Ontario government study argued that "there is a direct relation between the level of economic activity and the volume of highway travel," and then proceeded to set up a chain of cause and effect that rested on good highways. "Increasing industrial, agricultural and mining output, as well as expansion of commerce and tourism, are all promoted by highway transport. In turn, they generate new requirements for better highway facilities. The higher standard of living and more leisure made possible by expanding production are reflected in increased motor vehicle ownership and travel."¹⁹⁶ A Quebec report shortly after the war made the relationship even more stark. Highways, it said, are "of primordial importance for the future development of our Province."¹⁹⁷ R. O. Swain, head of the International Federation of Highways, saw even more far-reaching implications during a speech in Toronto in 1954. "The highways of tomorrow will provide an efficient and effective distribution of goods and commodities sufficient to generalize better living for peoples in all walks of life." It was not just goods however. "With the highways of tomorrow will come a better and closer relationship between Canadians of the East and the members of

your own family in Western Canada. Likewise, all of you will enhance life with a closer and friendly spirit of cooperation with your neighbors in the Western Hemisphere."¹⁹⁸

From a later perspective what is striking in the rhetoric of roads and cars between 1945 and 1965 or so is what might be termed the "absence of doubt." In more recent decades, as we shall see, the car's proliferation has generated heated debate. Environmental concerns, urban desires to preserve neighbourhoods, and a reaction against the free-wheeling materialism of earlier decades have created a counter-image to the powerful concepts of freedom and open road evoked by the car. The 1950s and early 1960s were not given to such discussion, however. This was an age when the concept of pollution was sufficiently undeveloped that architects designed "living garages," a sort of combination family room-garage. This way one could combine those modern symbols, the car and the TV, in one room!¹⁹⁹ The concern was not with constraining the vehicle but that highway construction keep up. One Minister of Public Works summed the mood up after complaining about the way in which car production had overwhelmed Canadian roads. "I do not contend that it is a bad thing to make motor cars; only that it is a bad thing that we cannot adequately serve motor cars with modern roads."²⁰⁰ It is a statement that summed up the road policy of two decades.

These post-war attitudes had a dramatic affect on government policy and the state of Canadian roads. In 1945 user taxes generated \$119.8 million from motor vehicles, considerably more than the \$73 million the provinces spent on roads. Fifteen years later a buoyant economy and large increase in registered vehicles had increased the user taxes to more than \$530 million. As Figure 19 shows, however, the provinces had increased expenditures even more, to \$657 million. Thus user taxes accounted for roughly 80 percent of expenditures.

A couple of trends are apparent in these figures. First, continuing a tendency that went back to the 1920s, the provinces were assuming an ever greater share of road costs. Provincial expenditures rose roughly nine times in the 15-year period after the war. In contrast, municipal expenditures had risen only 2.25 times. Second, from 1945 to 1960, transportation-related items were the single biggest cost of the provincial governments as a whole, representing about 22 percent of total expenditures. In 1975, by contrast, health, education, and social services all outranked transportation. Truly the 1950s were the age of infrastructure improvement.²⁰¹

Figure 19
PROVINCIAL ROAD EXPENDITURES, 1945-1960



Source: Derived from *Canada Year Book*, 1945-1963.

Aside from saying that this was a period of rapid development, it is difficult to describe in a coherent manner exactly what was done in the period from 1945 to 1965. One of the difficulties is that at mid-century there was a disparity among provinces in the number and quality of existing roads. Southern Ontario was already fairly well paved even before the war began. In contrast, when Newfoundland entered Confederation in 1949 it was still impossible to drive across the island and there were practically no paved roads outside of the immediate vicinity of St. John's. There were also extensive differences within provinces. British Columbia's lower mainland was also well served by roads, but much of the interior was impassable part of the year and served even in summer only by dirt or gravelled roads.

Still, allowing for some generalization, there are certain distinct patterns discernible.²⁰² This was a period of road improvement rather than one of new roads. The overall road mileage in Canada in 1965 was roughly the same as it had been two decades earlier. What changed was the quality of main and secondary roads as curves were straightened, surfaces upgraded and level crossings eliminated. There were also shifts of emphasis as time

went on. In the first years (roughly to the mid-1950s) the emphasis seems to have been on the increase of paved mileage. The provinces sought to extend the sort of roads that already existed around large cities to link smaller cities and towns. Paved mileage in Ontario, for example, increased by some 50 percent between 1945 and 1955. Other provinces, starting as they did from a lower base, had even more impressive percentage gains.²⁰³

By the late 1950s the emphasis shifted back to the links around and between the large cities. Congestion on the main arteries between large urban centres had become a serious problem. In response, planners and governments looked to multi-lane divided highways, or freeways. In 1945 the only inter-city divided highway in Canada was the Queen Elizabeth Way (QEW) between Toronto and Hamilton. It had opened in 1937. Even after six years of post-war construction there was only 202 miles (325 km) of freeway in the nation, practically all of it in Ontario. By the mid-1960s the system had expanded to 1,646 miles (2,649 km).²⁰⁴ Many of the highways most familiar to motorists throughout Canada were planned and constructed in the 10 years from the mid-fifties to the mid-sixties. These include Ontario's Highway 401, begun earlier but completed only in the early 1960s; the 403, linking the 401 to the QEW in Toronto; the Laurentian Autoroute, constructed in the late 1950s; the Sherbrooke Autoroute, constructed in the early 1960s; the twinned Highway 2 in Alberta between Edmonton and Calgary; the twinned highway eastward from Vancouver toward Hope; the Manitoba Ring Road and the Queensway in Ottawa.²⁰⁵ Nothing was more symbolic of the changes that had taken place since the war than the appearance across Canada of these limited-access, high-speed, divided highways.

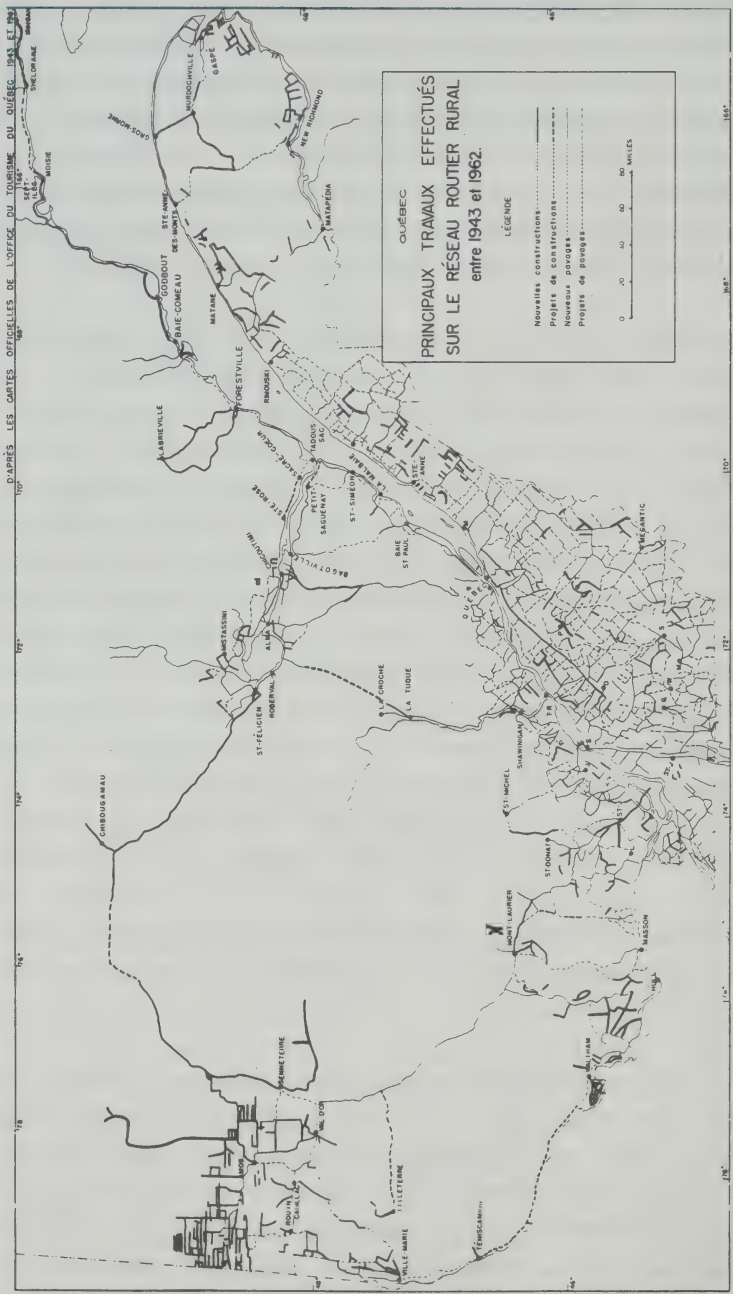
Though there is no "typical" province, the experience of Quebec in these years illustrates the type of activity that went on. Until the period after World War II, Quebec's efforts at road construction had been intermittent. There were several reasons for this. First, the number of car owners was much lower in Quebec than in Ontario. At the beginning of the war, for example, Quebec had 225,000 motor vehicles, compared with more than 700,000 in Ontario. Second, government philosophy emphasized the virtue of the parish and rural life and did not put a high premium on development that would link the countryside more closely to the town. Further, the same conservatism meant that government tended to move reluctantly into large-expenditure areas.

In 1944, however, a government commission investigated the state of Quebec roads. It found them abysmal outside the larger metropolitan areas and detrimental to trade, tourism and even to the prized rural life. There was an element of politics in this, as the Union Nationale party had just been restored to power and could blame its predecessors, the Liberals, for the ills that beset Quebec roads. Still, the report was essentially accurate. As of 1945 only about 10 percent of Quebec roads were paved, and only about 50 percent had any surface other than dirt. Likewise, only about 10 percent of roads outside of municipalities were kept open in winter. The great majority of road mileage was still seasonal, rough and inappropriate for the modern car.²⁰⁶

Such a situation was increasingly intolerable. The end of the war saw an upsurge of car ownership in Quebec, as elsewhere in the country. Even the valued rural farmer, as the government admitted, was increasingly dependent on good roads both for the marketing of goods and as a social amenity. As a result the government embarked on a program to reconstruct the roads of the province. In 1945 the budget for provincial roads was \$25 million. By 1955 it was more than \$93 million.²⁰⁷ The emphasis was initially on improving rural roads. Created partly as patronage, partly from philosophy and partly out of necessity, the government program nevertheless accomplished a great deal. As Figure 20, which is a reproduction of the original map, indicates, by the mid-1950s major improvements had been carried on through the Gaspé region, the Saguenay region and the Laurentians. Quebec could boast that almost 25 percent of its total road system was paved, the highest in the country.²⁰⁸ It was now possible to drive from New Brunswick to the Ontario border on modern, paved roads. Likewise, at least in the southern part of the province, most farmers were only a few miles at most from pavement. Equally dramatic improvements had been made in winter accessibility. By 1956 three quarters of the province's road system was open all year round.

By the late 1950s the original impetus of the 1944 plan was losing steam. The road system to the parishes was much improved, but, as with elsewhere in the nation, serious problems of congestion were developing around the major cities. Montreal's problems were especially acute. Traffic from the city northward to the Laurentians tripled between 1937 and 1954, and even though improvements had been made to the road system there, congestion was a major problem. In 1957 the government adopted a novel approach — or perhaps reinvented an old one — to resolve the growing

Figure 20
 MAJOR WORKS CARRIED OUT ON QUEBEC'S RURAL ROAD NETWORK BETWEEN 1943 AND 1962



Source: Michel Bérard, *Les Routes du Québec* (Quebec Department of Roads, 1964), p. 41.

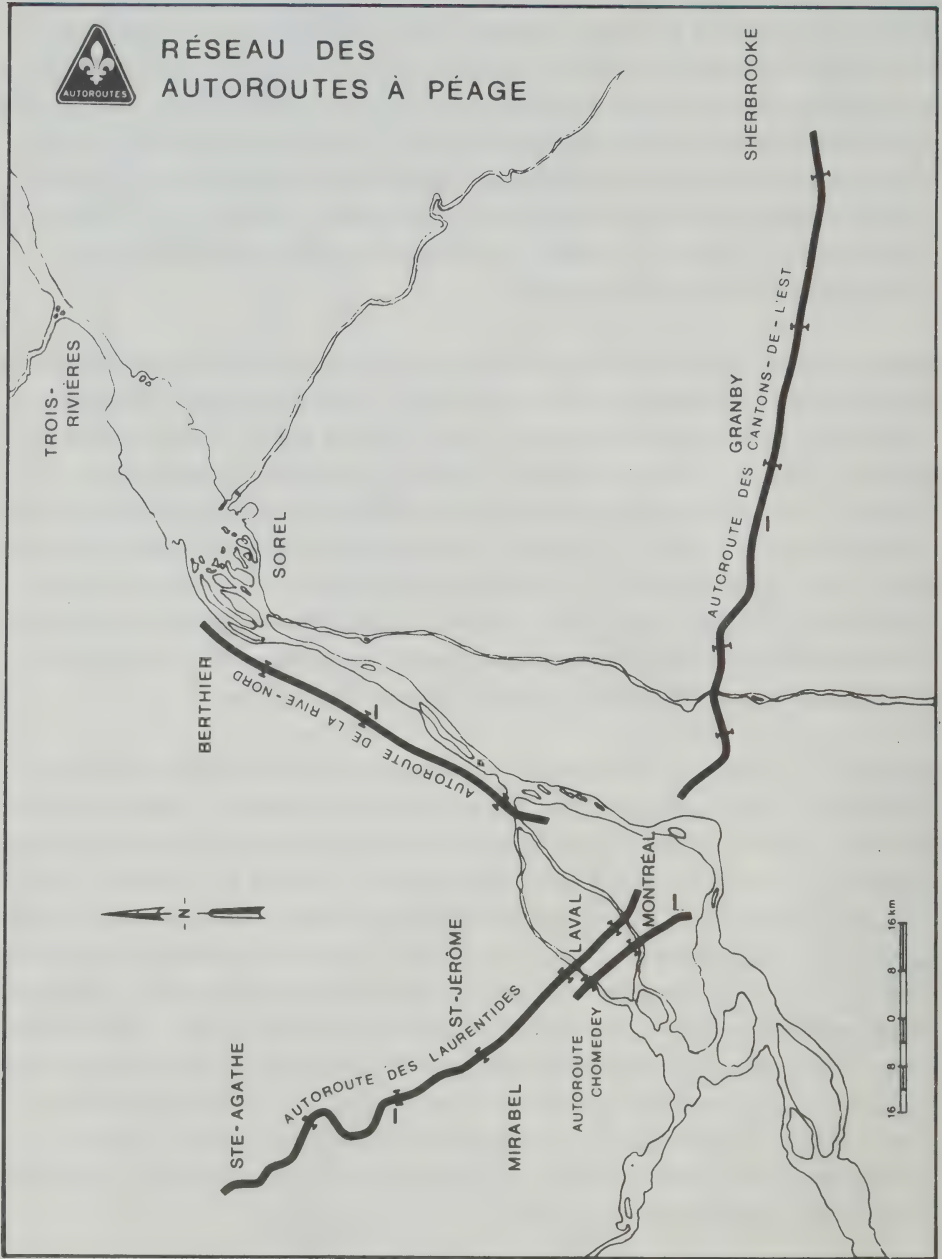
difficulty. The Laurentian Autoroute Authority (LAA) was formed. This was a private body that, with funds borrowed from the government, was to construct a modern, divided, six-lane highway from Montreal to Saint Jerome. LAA built the 39-mile (63 km) Autoroute by 1959 and operated it as a toll road. As with the 19th century toll road, the idea was that the public would pay for superior service, in this case a 70 mile per hour (110 km/h) uninterrupted trip to Montrealers' favourite holiday spot.²⁰⁹ With the completion of this route the LAA was given an expanded jurisdiction and began construction on highways south of the St. Lawrence. Eventually a complex of autoroutes became a basic part of Quebec's road system (see Figure 21 which is a reproduction of the original map).

Quebec nicely represents the problems of uniqueness and generalization in any discussion of roads. It was unique in its governmental philosophy and in its reliance on a semi-private road authority for some of its projects. At the same time it typified the general Canadian pattern outlined above. Initially the emphasis was on expanding the network of paved roads beyond the major centres. Then, by the late 1950s and early 1960s attention shifted back to the problems of major urban centres and the congestion that had become a part of life there. Also, like other jurisdictions, Quebec responded with the freeway solution. Multi-lane divided highways with high speeds and limited access seemed to provide the key to the future.

Whether in Quebec or in other provinces, the massive transformation of Canadian highways did not take place without problems. A massive allocation of government funds and national physical resources were dedicated to highway construction in the years after the war. As one speaker said to the Canadian Good Roads Association (CGRA), the 1950s were "the most frantic period yet" in the history of road building.²¹⁰ The problem was not a lack of money. This was a buoyant time for governments and the public seemingly had great enthusiasm for road construction. A survey taken in 1954, already a year of significant construction, asked if the public would be willing to pay higher taxes to generate additional construction. The answer was two to one in favour of the idea. One Highway Minister commented, "there is no government that is held back in its road and street programs by an unwilling public or by rebellious taxpayers."²¹¹

The problems came mainly from the sheer demand put upon physical and human resources. Provinces, especially smaller ones, had to build up the requisite staff of experts to handle the complex engineering issues of modern

Figure 21
NETWORK OF TOLL ROADS, QUEBEC



Source: Map reproduced from Office des Autoroutes du Québec, *Rapport d'activités*, 1981, p. 5.

road-building. This was a time when engineers were in short supply, however, and it was often very difficult to get or keep skilled and experienced road engineers. The same was true of skilled labour. In addition, during the the Korean War (1950–53), the Government imposed allocation rules for vital goods such as steel. As a result many provinces had to delay bridge-building programs considerably.²¹² The provinces desperately tried to keep up to public demand. It is hardly surprising, therefore, that pressures began to develop for the Dominion government to take an active role in highway development.

THE HIGHWAY CRISIS AND THE FEDERAL PRESENCE: 1949–1967

Until the 1950s, federal involvement in road-building was sporadic. The various Canada Highway Acts of 1919 to 1928 had allocated \$20 million to assist provinces in the years after World War I. In the 1930s roughly \$30 million had been spent on a road program designed to assist the unemployed. During and after World War II, however, the federal government had withdrawn from most road construction. From 1946 to 1949 federal highway expenditures averaged less than four percent of that of the provinces. Indeed, cooperation between provinces had been much more significant than between federal and provincial governments. For one thing the CGRA had evolved into a quasi-interprovincial forum on highway policy. Provincial ministers of highways or their equivalents routinely headed the organization, and provincial funding supported it. The Good Roads Technical Council provided a liaison branch on highway specifications, construction methods and surfacing materials. There was also direct interprovincial cooperation on the necessary details of highway construction. A road to the provincial border was quite useless, after all, if it didn't connect with another road.

By the early 1950s the provinces were clamouring for assistance. Several arguments were used to justify the demands upon the federal government. First, roads on a modern scale had not been envisaged by the framers of the Constitution, and assistance was therefore needed. Second, the federal government reaped various fiscal benefits from the improved road system. Not only were there excise taxes on cars and, during the war, federal gasoline taxes, but the commerce that generated federal taxes depended on the road system being built by the provinces. As A. R. Morrison, President of the Canadian Automobile Association, said at one point, road construction has always been based on a user-pay philosophy. "Those who benefit from roads should pay for them." The Dominion benefitted and the Dominion

should pay.²¹³ Third, there were even attempts during this Cold War period to borrow American precedents and argue that a good road system was key to national and civil defence. "Should atomic warfare break out, there is only one defence for civilian population — flight. Transportation itself is not the problem. The entire population of our cities could be moved if there were roads to carry the vehicles."²¹⁴ It is doubtful that any specific argument carried much weight. Their urgent repetition, however, did convey the sense of crisis felt by the provinces as well as the message that there would be little complaint about federal invasion of provincial jurisdictions.

The post-war years did bring greater federal involvement in road construction. At the end of World War II there was a strong inclination on the part of federal ministers and officials toward central planning. The Rowell-Sirois report still echoed in the mandarin's ears, and at the 1945 Dominion-Provincial Conference on Reconstruction numerous plans were put forward to implement new programs. In a very practical sense, the federal government had the money. Despite the debt created by the war, post-war revenues were healthy. The Government ran surpluses nine out of the 10 years between 1947 and 1957.²¹⁵ Available money allowed the officials of the day to follow their inclinations! In other words, the general climate at the federal level was right for a positive response to provincial concerns.

The highway issue also had a specific appeal for the federal government. The Canadian highway system was, as has been mentioned, regionally oriented. This was understandable given that it was the provinces that were doing the planning. For a long time, however, car enthusiasts, road planners and federal officials had lamented the absence of a meaningful cross-Canada highway. By the 1950s it was theoretically possible, at least in summer, to make the trek from coast to coast. It was not practical, however. The east-west connections were generally weak and in parts of the country non-existent. Anyone travelling from east to west invariably headed south of the border. The dream of a national east-west highway, however, was a natural one for a generation that was nationalist and inclined toward big developmental projects.

The Trans-Canada was first mentioned in its new form at the 1945 Dominion-Provincial Conference on Reconstruction as one of many schemes for the post-war period.²¹⁶ Unlike many other ideas at the conference, though, it was not killed by provincial opposition. In 1948 a more specific proposition received support from most provinces, and in 1949 Robert Winters, the

Minister of Reconstruction, presented a bill to Parliament.²¹⁷ With support from all sides of the House, the *Trans-Canada Highway Act* became law in December 1949.

The parliamentary discussion on the bill is revealing. This was a project in which national pride and symbolism rather than commerce dominated — national unity, trade, public demand and national defence were all given as good reasons for the immediate pursuit of the highway project. Though it would never command the federal resources of its 19th century counterpart, there was a sense that the road was the 20th century equivalent of CPR. “The fathers of confederation,” said one MP, “realized that the building of a railroad was a national responsibility. I believe the government should realize that the building of a trans-Canada highway is also a national responsibility.”²¹⁸ This was echoed later by the minister in charge of the project. “Our undertaking is the largest East-West construction project since the building of our Canadian Pacific and Canadian National Railways.”²¹⁹

The bill itself was straightforward. As with many other federal-provincial programs, there was a cost-sharing arrangement. The federal government agreed to pay 50 percent of the costs on approved highway projects. The Department of Reconstruction was initially the federal agency involved, but when Robert Winters moved over to Public Works he took the Trans-Canada project with him. To be approved the project had to fit into the overall Trans-Canada route. Construction would have to meet specific standards set by the federal government after negotiation with the provinces. The federal government would have no proprietary interest in the highway after it was built and no responsibility for maintenance.²²⁰ The maximum federal contribution under the bill was to be \$150 million, and the completion date was set for December 1956. By the end of 1950 all provinces except Quebec and Nova Scotia had signed the agreement; Nova Scotia signed it in 1952; Quebec, in 1960. In the meantime, though, Quebec built its east-west provincial highway to the general standards set out in the Trans-Canada specifications, so Quebec’s delay in joining did not really affect the progress of a Trans-Canada system.²²¹

The possible rate of construction as stated in the agreement was unduly optimistic. Given the already extensive highway program in most provinces, it was impossible to shift sufficient resources to the Trans-Canada in time to complete it by 1956. Provinces dedicated resources and effort to the Trans-Canada only in proportion to the way it fit provincial priorities. Thus, for

example, Saskatchewan and Manitoba saw the development of a first-class east-west highway link as highly desirable and made considerable progress. For Newfoundland the highway and the federal funds provided a much desired opportunity to have paved highway communication across the entire island, but limited local funds and the poor state of the provincial highway system made it impossible to move too quickly. On the other hand, Ontario was interested mainly in the Trans-Canada in the south and did not devote effort or attention to the difficult stretch around the Great Lakes. British Columbia faced a horrendous task through the mountains and proceeded slowly. Finally, even had the will been there, the amount of money and time needed were greater than original projections had suggested.

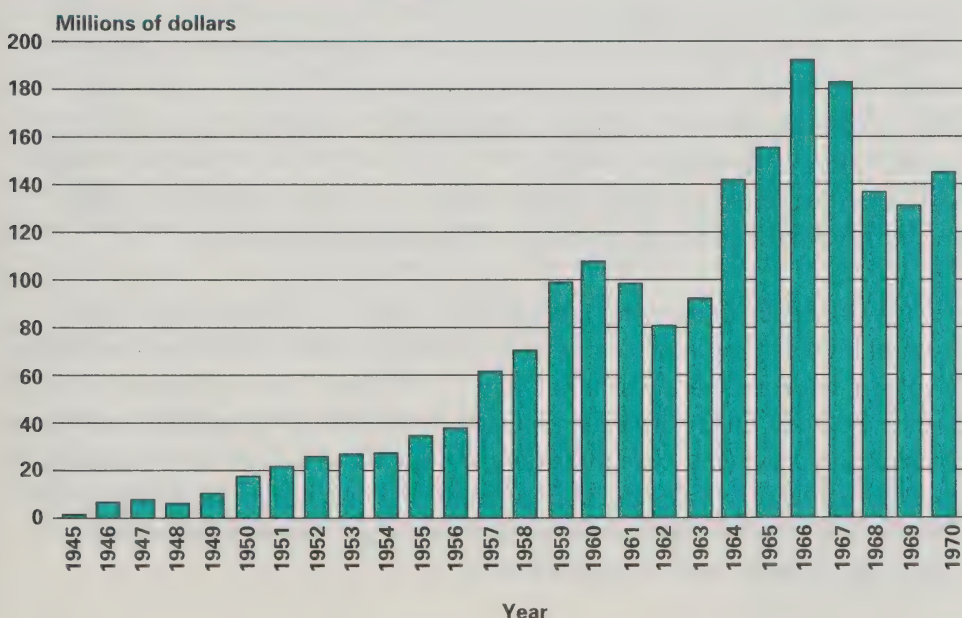
After early brave pronouncements on the progress of the highway the federal officials soon became worried. Unless the provinces "markedly step up their efforts," warned the Department of Public Works in 1954, very few would complete the work by the deadline.²²² By 1955 the "very few" provinces had become none, and by October of that year a federal-provincial conference was called to amend the agreement. It was a sign of the fiscal health of the federal government and its commitment to the project that it not only extended the deadline for completion to 1960 and increased the amount of money available but changed its funding formula as well. Robert Winters, the Minister of Public Works, noted that there was real lack of progress in many of the provinces on about 10 percent of the mileage. In most cases this was the most difficult or lowest priority section of highway — around the north of Lake Superior in Ontario, for example. In response Winters created what became known as the 90/10 formula: the federal government would pay 90 percent of the cost on 10 percent of the mileage within any given province.²²³

This broke the logjam. All sections were now tackled and real progress was made even on the most difficult sections. Still, it was expensive and time consuming. The federal government had to increase its contribution to \$350 million in 1959 and to \$400 million in 1960. It extended the deadline time and time again. In 1962 it held a lavish and much publicized opening ceremony at Rogers Pass.²²⁴ In fact, though, this was public relations more than reality. It wasn't until the late 1960s and several million dollars more had been spent that the Trans-Canada Highway was fully completed. Still, by the time of the Rogers Pass ceremony it had finally become possible to drive through Canada on "passable mileage" from sea to sea.²²⁵

There were other federal road projects. With the Trans-Canada still under way, the new Conservative government of John Diefenbaker began the “Roads to Resources” program in 1958. The fiscal arrangements were the same as with the original Trans-Canada proposal: a 50 percent cost-sharing arrangement with the provinces. In a sense, however, this program was a precedent in its own right. Unlike the Trans-Canada project, which was focussed on a particular route and a particular nationalist goal, this program was open ended in terms of roads and brought the federal government into a project that was potentially open ended in terms of commitment to provincial highway programs along the frontiers.²²⁶ It is also a prime example of what was earlier termed the absence of doubts. Neither government nor opposition nor interest group raised the sort of environmental and sociological questions that such northern projects would create in later decades.²²⁷ Thus for the next 10 years the Roads to Resources program poured another \$75 million dollars into provincial road-building schemes.²²⁸

The Trans-Canada project and the Roads to Resources program seemed at the time to indicate that the federal government had committed itself to a significant ongoing presence in road-building. As Figures 22 and 23 show,

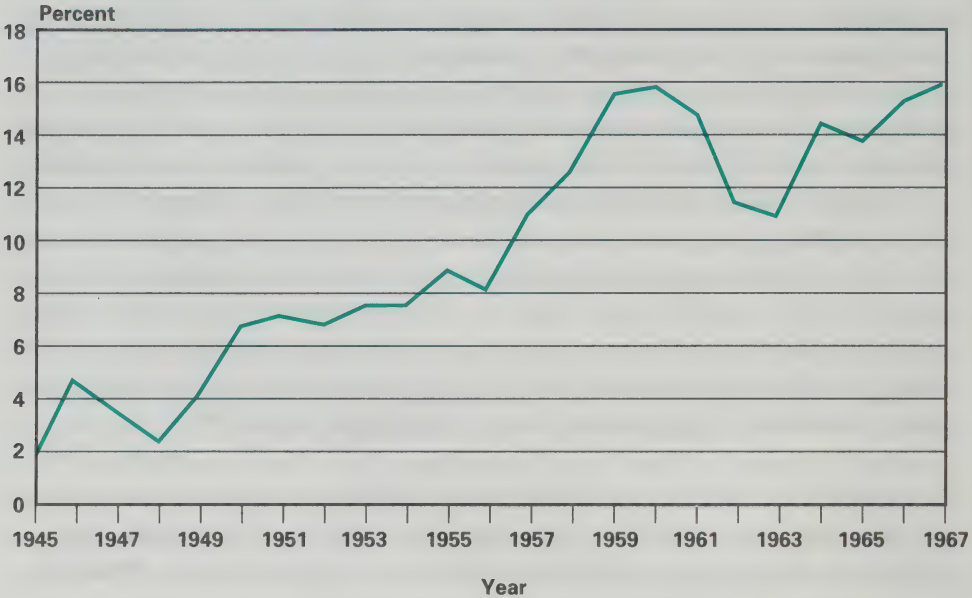
Figure 22
FEDERAL EXPENDITURES ON HIGHWAYS, 1945–1967



Source: Statistics Canada, *Road and Street Mileage and Expenditure*, Catalogue No. 53-201, for years 1945–1970.

Figure 23

THE GROWING FEDERAL CONTRIBUTION TO HIGHWAY COSTS: FEDERAL EXPENDITURES AS A PERCENT OF PROVINCIAL



Source: Statistics Canada, *Road and Street Mileage and Expenditure*, Catalogue No. 53-201, for years 1945–1968.

the contributions of the federal government to roads had been going up steadily in both absolute and relative terms through the 1950s and early 1960s. The relative figures are especially significant, given the fact that provincial expenditures had been increasing at a rapid rate throughout this period. More and more it looked as if the pattern of centralization that had seen road expenditures move from locality to province was continuing, this time with the growing presence of the federal government.

Yet the principles underlying federal involvement in roads were vague. Federal involvement really came about because there was the money to get involved, because there was a hue and cry for help from the provinces, and because certain major projects appealed to federal politicians and planners. Though the provinces were increasingly dependent on a federal presence in road construction, there was nothing here that was not more or less ad hoc. Should any or all of the above conditions change, federal involvement could easily shrink back to minor levels.

THE AIRPLANE

Airline travel as a mass phenomenon is a very recent event. As of 1962 nearly 60 percent of adult Canadians had *never* been on an airplane. Even in 1976 a full third of Canadian adults had never flown.²²⁹ Given these figures and given the high percentage of people who travelled only rarely, it would seem that even after 20 or 30 years of extremely rapid expansion the airplane was a form of transportation reserved under normal circumstances for the well-off traveller and for the business executive whose company saw the extra costs as worthwhile in terms of time saved. Nevertheless, rapid growth in the 20 years after the war provided a new source of competition for the railways. They had already lost much of their local and feeder traffic to the car. Now, at the other end of the market, the airplane was an ever more potent alternative.

Commercial airline service on a small scale dates back to the 1920s in Canada. Bush pilots ferried customers to locations inaccessible by railway or car. Municipalities developed rudimentary airstrips to encourage the development of local flight. In the 1930s business interests, including the railways, attempted to develop a national airline under the title of Canadian Airways Limited. Government had an interest in all of this preliminary activity and wanted to encourage some sort of coherent domestic system that had a chance of survival. Given this and given the dominance of the railways in passenger transportation, it made sense to organize air travel as an extension of the two giant railway concerns. The idea was that CNR and CPR would put up the capital and create something like a reorganized Canadian Airlines. That failed because CPR had no desire to move into a corporate structure involving a large Government presence. By this time C. D. Howe, the Minister of Transport, was enthused by the idea of a national airline, however, and he pushed forward an organization in which the new national airline, Trans-Canada Air Lines (TCA), would be a government body, wholly owned by Canadian National Railways.²³⁰

There is a large element of the ad hoc in the way this all unfolded. The Government decision to create a new nationally owned transportation corporation at a time when CNR losses were seriously affecting the national debt seems curious indeed. It can be explained only by the fact that initially

there was supposed to be a large element of private capital; by the fact that the Minister of Transport, C. D. Howe, was an enthusiast for such schemes; and by the widespread assumption that the airline business was, after all, a fairly small-scale enterprise. Before the war, TCA's operating expenses were less than 3 percent of Canadian National's. During the war TCA slowly expanded operations to meet demand. Not surprisingly, such expansion was severely limited by the demands the war placed upon the economy as a whole.

Thus, at the end of the war, commercial air travel was still in its infancy. Indeed, from the perspective of the 1990s, the growth of air travel after the war dramatically emphasizes the technological and commercial revolution that has occurred in the past 45 years. As of December 1945, TCA's total fleet consisted of 11 Lockheed 14-08s, 14 Lockheed Lodestars and three DC3s. Total seating capacity for the entire fleet was 369 people! The airline carried 183,000 passengers, or one third of one percent of the number of people the railways carried that year. TCA had only 5,299 route miles (8,528 km), and several major Canadian cities were not yet served by air at all.²³¹

There were good reasons for this primitive service, aside from the disruptive effects of the war. Airline travel was fast, of course, but it was still relatively uncomfortable and unreliable. None of the aircraft commercially licensed in Canada were pressurized. This meant that oxygen masks had to be broken out for flights over the Rockies. These limitations, coupled with the harsh vibration of the twin-engine piston propellers, created a real necessity for that airsickness bag (which is still there as an anachronism in modern flight). Weather could prove a serious problem for even the most hardy. Because aircraft could not climb above weather, they were often grounded by it. Longer flights, especially in winter, often arrived many hours, or even days late, if they left at all. The public responded by avoiding the plane in the winter season, contributing to an amazing 44 percent differential between summer and winter traffic in the years immediately after the war.²³² Even when there were no problems or delays, longer flights were hardly speedy by today's standards as the following morning transcontinental flight schedule for 1945 indicates.

Trans-Canada Air Lines
1945-46 Flight Schedule
Montreal to Vancouver

8:25 a.m.	Depart Montreal
9:10 a.m.	Arrive Ottawa
10:55 a.m.	Arrive Toronto
2:10 p.m.	Arrive Kapuskasing
4:20 p.m.	Arrive Armstrong
5:55 p.m.	Arrive Winnipeg (CST)
7:20 p.m.	Arrive Regina
9:55 p.m.	Arrive Lethbridge (MST)
12:00 a.m.	Arrive Vancouver (PST)

Source: National Archives of Canada, Record Group 46, Series DIII, Vol. 656.

In the years immediately following World War II, however, the airline companies moved aggressively. Both TCA and Canadian Pacific Airlines (CPA) made large-scale purchases of new equipment, though Government restrictions meant that CPA's mandate was primarily international rather than domestic. The Canadair Northstar was introduced in the late 1940s in an arrangement, partly subsidized by Government, to provide the aircraft company with peacetime work. The Northstar was faster and larger than earlier planes, with a seating capacity of 40. It was also pressurized. This allowed easier flights over the Rocky Mountains and better escape from weather patterns. The duration of flights from Montreal to Vancouver decreased from 16 hours to 13 hours. The Northstars also allowed the introduction of commercial trans-Atlantic flights, with planes travelling to London from Montreal in 15 hours by 1947.²³³

The Northstars were followed through the 1950s by more advanced planes. The 60-seat Super Constellation was first delivered to TCA in 1954. It was the largest commercial plane ever licensed in Canada and was such a novelty that some 30,000 Canadians lined up at airports when the airline put it on public display. It was followed by the Viscount (1955) and the Vanguard (1960). All of these planes represented significant improvements in comfort, cost-effectiveness and speed. Yet all were introduced within 15 years of the end of the war. In 1961 the arrival of the first DC-8s brought commercial jet travel to Canada.²³⁴

The investment by the airlines was supported by growing Department of Transport investment in facilities. Navigational supports, airstrips, terminals and an improved meteorological service all developed rapidly to match the changing technology and growing demands of the airplane. As Table 4 indicates, Department of Transport capital expenditures on air services multiplied nearly ninefold in the decade 1951 to 1961. By the end of that time there were 120 airports operated by the Department.²³⁵ It was a level of subsidization that caused bitter resentment among railways.

Table 4
TRANSPORT AIR SERVICE CAPITAL EXPENDITURES

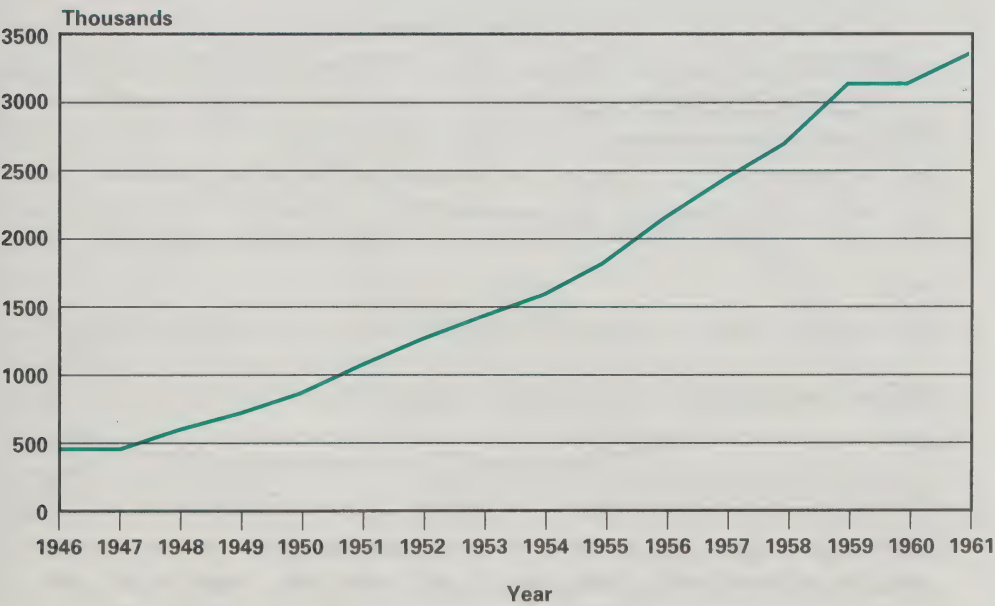
Year	(\$000)
1951	8,625
1952	9,896
1953	12,141
1954	14,742
1955	25,625
1956	32,952
1957	46,018
1958	63,003
1959	64,555
1960	63,420
1961	73,534

Source: Canada, Department of Transport Air Services, *Canada in the Jet Age, A Report on a Study of Department of Transport Air Services Needs, 1962-1972*, November 1962, Appendix.

The investment in equipment and airports helped create the most important statistic of all, a tremendous increase in passenger travel. Each year after the war there were new customers. TCA's 183,000 passengers at the end of the war increased to 427,967 by 1947, even before the Northstar was fully in service. By 1949 this number had increased to 648,574. By 1954 it was 1,438,349, an eightfold increase in a decade. Route miles of the airline increased from 5,299 (8,528 km) at the end of the war to 8,362 (13,457 km) in 1950. By then most major and intermediate Canadian cities had scheduled air service. Henceforth airline expansion was concentrated domestically on increased frequency rather than on new routes. The growth of international connections across the Pacific (CPA) and to the West Indies (TCA), however, expanded international mileage.²³⁶

As these figures imply and as Figure 24 further demonstrates, the growth in air travel was tremendous in the 15 years after the war. At the war's end, air travel was still an uncommon experience for Canadians. Senior business and government officials used it under the pressure of time. Tourism was developing, but it was still the rare family that in 1948 or, for that matter, 1953 would pack the children into the Northstar and head off across country. It required a sense of daring, above average-affluence and probably distant relatives before people thought of travelling as families on the plane. Most Canadians had never been on an airplane. By the end of the 1950s that was changing. Air travel was not yet the practically universal phenomenon it became by the 1970s. Still, it was by 1960 a common form of business travel and was seriously affecting the railway in that market. Tourism was expanding as well: families were lured by the introduction of holiday packages, southern routes and more comfortable aircraft.

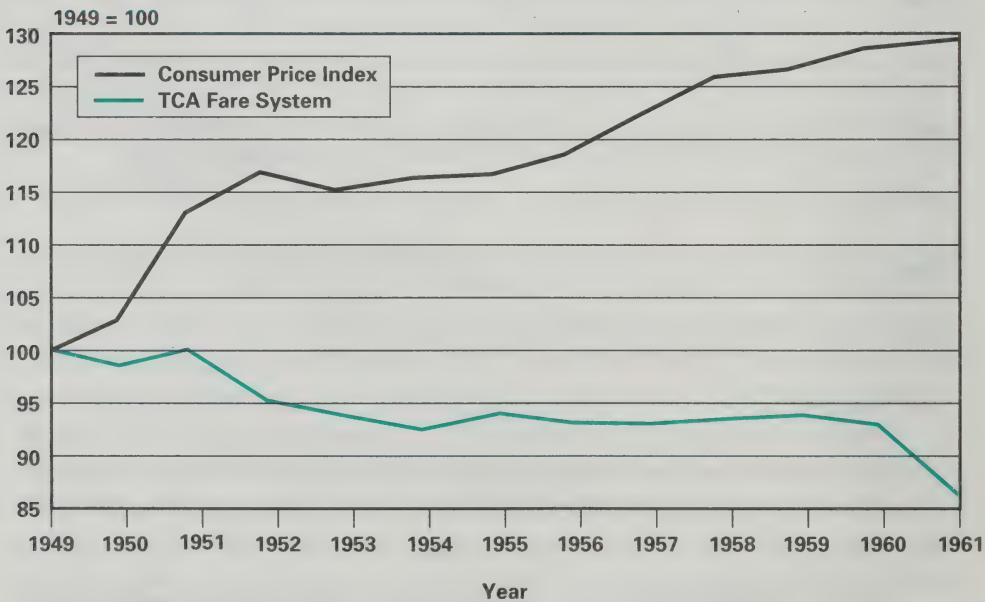
Figure 24
NUMBER OF AIRLINE PASSENGERS



Source: Derived from Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series T199.

Contributing to this growth, and resulting from it as well, was the rapidly declining real cost of air travel. Until well into the 1950s air fares were quite high and thus reserved for those with expense accounts or unusually high incomes. Air travel might have been faster than the train, but it was also considerably more expensive.²³⁷ Throughout the 1950s those costs kept coming down. Economies of scale generated by larger airplanes and increasing passenger volumes meant that airlines could absorb increases in the cost of living, whereas railways could not. Thus in 1954 TCA could boast that “there has been no major fare increase by TCA since 1947” and that “It is the Company’s aim, within the bounds of economic stability, to make air travel financially available to more people.”²³⁸ By the 1960s the introduction of tourist and then economy fares lowered the cost of air travel still further. Now TCA could boast that Canadians had “the lowest air fare structure in the world” and could back it up with a chart that showed the plummeting real cost of fares since the war (see Figure 25).²³⁹

Figure 25
INDEX OF TCA FARES VERSUS CONSUMER PRICE INDEX, 1949-1961



Source: Trans-Canada Air Lines, *Annual Report*, 1961.

The perception of the airplane and air travel during the years of growth was complex, though firm conclusions await full-length study. This perception combined a number of ideas, some of them contradictory. The airplane represented adventure. For most individuals the idea of airline travel was still exciting and a little daunting. Even after the Viscounts, Vanguards and especially the DC-8s made travel comfortable and routine, the memory of the earlier noisy planes, unpressurized cabins and vulnerability to weather remained in the public mind. There is also the simple fact that the idea of going up in the air was anything but routine when the majority of the populace had never done so.

This was not the image airlines wanted, and through the 1940s and 1950s their ads downplayed adventure. In fact, the advertisements downplayed the entire experience of travel. Instead they emphasized one of two things. The first approach was to stress the technological advantages of air travel. Advertisements talked about the newness, superiority and design of the latest airplanes. The second, and more common, approach was to ignore the travel experience and technology altogether in favour of emphasizing the point of destination. The message was significant both for what it conveyed and for what it ignored. The airline could take you almost instantly to a reunion with friends, loved ones or tourist spots far away. For the most part, however, they did not really pretend they would do so in great comfort or that the actual act of travelling would be a highlight of your vacation or business travel. The romance of the airline, then, was not in the mode of travel but in the destination.²⁴⁰

The message got through. By the 1960s the miniscule competitor of the 1940s had become a serious challenger to railways. There was now one airline passenger for every five rail passengers. Moreover, those airline passengers were being drawn from two very important market segments. First, the airlines were pulling away the long-distance traveller, the one least likely to turn to the car as an alternative. Second, they were drawing the affluent business traveller, the regular user both of the train and of luxury services like sleeping and dining cars. The car had done much to destroy the train's hold over the short- and medium-haul passenger. The airline prevented the railway from holding control of the long-haul market. Increasingly the railway, which had possessed a near-monopoly on land travel only a generation

before, was without a certain market for its product. This meant that the passenger business, never a very profitable aspect of railways, was beginning to stand out as a very expensive activity.

Even so, airline expansion has to be kept in perspective. It would take another major era of growth before the airplane overtook the train as the most common form of intercity public transit. By the early 1960s the first great flurry of growth in air travel was over. The airlines had reached the limits of their current market, and the phenomenal increases of the 1950s slowed down to staid growth rates of 3–4 percent a year through the recessionary early 1960s. This slowdown came just as the airlines were investing heavily in jets, and the volume-based profits of past years gave way to operating losses. The expenditures, the slow rates of increase in passenger numbers and the losses all meant that the airlines had to pause to take stock through the first part of the decade. The airlines were discovering other factors affecting their ability to cover the travel market. The growing size of airplanes, the development of jet service, and the loss of passengers to continually improved cars and roads made short-haul flights to regional centres unprofitable. Technology also forced change. By the late 1950s the new, larger aircraft could not land on the runways of some of the smaller airports. As a result TCA began, for the first time, to abandon centres. In 1961, for example, it made moves to shut down routes to a number of Manitoba and Saskatchewan communities.²⁴¹ By the 1960s, therefore, the first great phase of airline expansion was over. For at least the next several years the inroads the airlines made into the intercity market were modest. Not until the 1970s did the next era of significant expansion begin. By that time the railroads were already beginning to retreat from the passenger business.

There is a final point to be made about the development of air travel and that is that there was little effort to develop an integrated train-plane system. From the beginning the planes competed with rather than extended the railway network. There is a certain irony in this, given that initially both TCA and CPA were subsidiaries of railway corporations. In part this was a reflection of Government attitudes. As Studnicki-Gizbert has noted, transportation policy in Canada was “modally oriented” and often seemed designed to keep the various services separated rather than integrated.²⁴² In part this may have reflected the traditional view that railways were too powerful and needed control rather than extension. At any rate, despite the existence of the Department of Transport and despite the nominal ownership of TCA by

CNR, airports and air services developed apart from rail services. Even the partial exception of CP's fleet of ships, trains and airplanes made little difference to the outcome. The ships were being sold off, and the passenger rail service was declining in frequency and importance.

THE POLITICIZATION OF THE RAILWAY PASSENGER ISSUE

Perhaps surprisingly, railway passenger service was not a particularly sensitive issue until well after World War II. Of course people had always wanted railways to run in their communities or regions, and the assumption had been that such lines included adequate passenger services. There was also a recognition by the railways that passenger services of high quality were expected. If the railways forgot, they were quickly reminded. Thus when CNR president Donald Gordon considered reducing transcontinental service after the war, he backed off because it hurt the public image of the railway to even consider such cuts. "Its impact on our freight service was much greater than expected. We saw that our analysis of what might be called the psychological effects of our cancellation had not been good."²⁴³ Nevertheless, until the 1950s neither the railways nor Government focussed on passenger traffic policy. It was an issue largely buried within the more contentious problem of freight. Once the question of passenger rail did emerge as a distinct issue, however, it quickly became politicized. For underlying it were not only economics and local community service but also highly volatile regional and nationalist sensitivities. These sensitivities shaped the railway and government response to the new competitive era that existed and led directly to the current policies and problems of passenger rail.

The relationship between passenger service and regional grievance is especially important in understanding what has happened since 1950. Of course, regional concerns have always played a part in railway policy. As we saw earlier, the Canadian government gave support to the Intercolonial, the Prince Edward Island Railway and CPR in response to regional concerns. Even after the railways were completed, regional politics often seemed to be railway politics. Yet, except for the past few decades, passenger service was not a part of this debate. Only as that service went into decline did the regions of Canada make the issue of adequate passenger service a matter of regional equity. Once they took up the issue, however, they could draw upon a rich tradition of regional rhetoric about the proper role of railways.

They could also point to policy precedents in which railway activities had been constrained in the name of regional concerns. That tradition involved freight rates; accordingly, discussion of passenger transportation and regional grievance must begin not with passengers at all but with freight.²⁴⁴

Both the Maritimes and the West have been very much involved in the regional criticisms of the railway system and of Government policies toward railways. The precise timing and nature of their concerns have been different, however. In the West many of the battles were fought and resolved before the turn of the century. From the beginning there was a tremendous dislike of the monopoly position of CPR. The lack of competition allowed the railway to make arbitrary rules on, say, loading and unloading grain unless the Government intervened. It also enabled the railways to set rates much higher in the West than in the East. In the 1880s, for example, general freight rates in Manitoba were about 50 percent higher than in Ontario. There were also a great many complaints about the absence of branch lines. CPR's policy in this area reflected a sensible, even cautious, policy of building only as traffic warranted. To the pioneer settler or local business on the frontier, however, the railway was the lifeline to the outside world, and it was natural that they would press for more rapid extension of the railway network.

By the beginning of the 20th century the West had achieved many results from its continual political agitation. The Crowsnest Pass agreement in 1897 set fixed rates (the Crow rates) for the hauling of grain and eased Western concerns in that crucial area. At the same time, the government further recognized the national importance of the relationship between the railways and the wheat economy by passing such legislation as the *Manitoba Grain Act*. Then, of course, the new transcontinental railways broke the monopoly of CPR and increased the area of the region easily reached by railway. By the 1920s the West was relatively well covered by branch lines, and though everybody loved to hate CPR, Western grievances had turned to other issues.²⁴⁵

In the Maritimes the timing was different. Discontent with the fruits of Confederation had created an on-again off-again series of complaints about regional issues. That discontent, however, had abated somewhat in the prosperity of the Laurier years and, at any rate, had never been focussed on the Intercolonial Railway in particular. By the 1920s, however, the grievance was much worse. The Maritime economy was shrinking relative to the rest

of Canada, and while other parts of the country enjoyed a reasonable prosperity by 1923, the Maritimes remained mired in recession. The situation had deteriorated so badly in both political and economic terms that a royal commission (the Duncan Commission) was appointed to investigate Maritime grievances.²⁴⁶

By now the Intercolonial Railway was very much a focus for Maritime discontent. This was due partly to the animosity generated by freight rate changes shortly before World War I and partly to the absorption of the railway into the new CNR system. There was more to it than that, however, for the railway was a symbol of what Confederation had promised to the region and not delivered. This view, which remained fairly constant, was neatly summed up by a Nova Scotian submission to the Duncan Commission.²⁴⁷ Whereas, the presentation concluded, the Province of Canada had numerous reasons to support Confederation, Nova Scotia had only two: "1. That an Intercolonial railway be built." and "2. That in general, and specifically as a result of the building of that railway, these provinces were to share in the internal commerce which the creation of the Dominion was intended to promote." As the trade was not up to expectation and as, indeed, the Maritimes had declined economically since Confederation, the region was entitled to further support on "equitable and national grounds."²⁴⁸

The claim is an interesting one, for it ties the fate of the railway and the charges by a railway not to a specific legal argument but to an historical intention or assumption. It was no longer sufficient that the railway was built and operated; it should be built and operated in such a manner as to overcome regional disabilities. The concept of the railway as a political instrument of regional equalization is clearly expressed here. Moreover, the principle was accepted, at least implicitly. In the wake of the royal commission the Dominion government passed the *Maritime Freight Rates Act*, which directly subsidized Maritime railway shipments as a part of the national arrangement.²⁴⁹

The Crow rates and the subsidy clauses in the *Maritime Freight Rates Act* therefore were precedents for treating railway charges not as instruments of profit and loss but as national policy instruments in the face of regional concerns. Nor was there any doubt about the right of the the Board of Railway Commissioners to determine both freight and passenger charges. In so doing they took "equitable treatment" and other such issues under

consideration. Generally, however, the Board did not see its mandate as extending in a very broad way to issues of regional equalization. Nor did it see that as something for which the railways were responsible. "Railways are not required by law," it wrote in one decision, "and cannot in justice be required, to equalize natural disadvantages such as location, cost of production and the like."²⁵⁰ There were thus two somewhat antagonistic principles at work within the governmental and public perceptions of the railway's role. All agreed that equitable rates were desirable, and this was worked towards throughout the 20th century. At the political level, however, and especially in the East and the West, railway rates were seen as instruments of a broad national policy that transcended the interests or rights of the individual corporation.

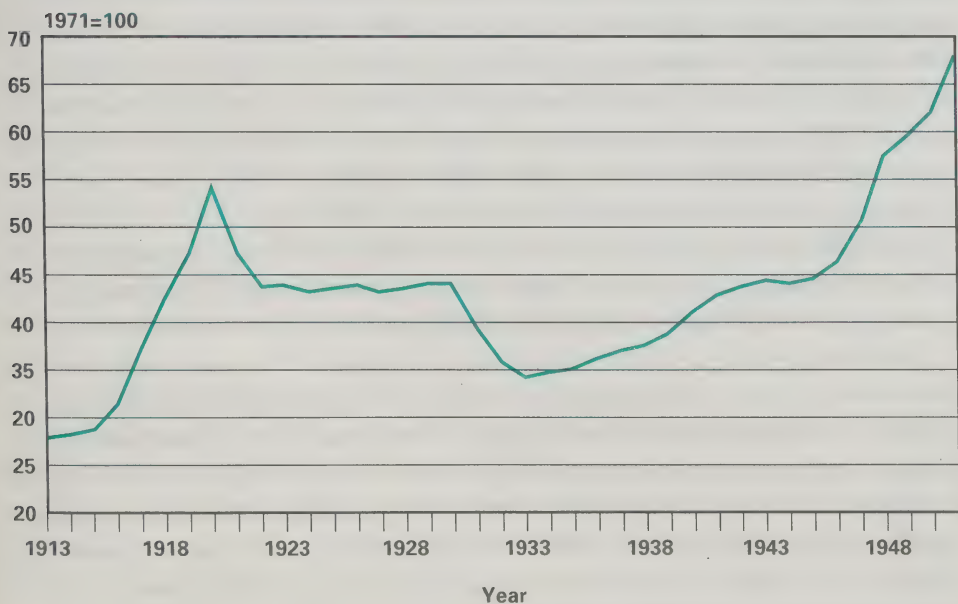
After the Duncan Commission the whole issue of freight and passenger rates became less controversial. This was due in large part to the simple fact that rates were not increasing. In the wake of the inflation of World War I there had been a series of rate increases approved by the Board of Transport Commissioners. Thereafter, however, the regulated passenger and general freight rates remained unaltered. (The first-class passenger rate was 3.45 cents per mile except in the mountains, where it was 4 cents a mile.)²⁵¹ The maximum proved profitable for the railways as the general price level declined after 1920. During the Great Depression, of course, railways found they had to lower rates much below the maximum allowable in order to generate business. During World War II, as we have seen, wage and price controls were in place. Besides, the railways made a healthy profit within the existing standards. Thus, though there was a regulatory system, the rate levels had been set so generously in the 1920s that in effect market rates, not regulation, determined passenger and freight tariffs between 1921 and 1945.

After World War II all that changed. As rates again became unstable they again became more political. As wartime price controls came off, inflation increased and for the first time the price index rose above what it had been just after World War I. The Korean War, which began in 1950, added to inflation. As Figure 26 indicates, for the first time since the years immediately after World War I there was a significant increase in the cost of living in Canada. Two additional factors made this increase in prices worse for the railways. First, there was a growing discontent in the workforce. In the immediate post-war years a series of labour disputes disrupted railway traffic and increased labour costs. Second, the railways were faced with daunting

capital costs. Deferral of capital purchases during the Great Depression now came back to haunt the companies. In addition new technologies, such as efficient diesel locomotives, appeared necessary if the companies were to operate efficiently. The conversion process, however, would be an expensive one.

The old standard rates quickly came under pressure in such circumstances, and the Board of Transport Commissioners was faced with a series of requests for adjustment. There were freight rate adjustments of 21% in 1948, 8% and then a further 16% in 1949, and 20% in 1950.²⁵² The increases reawakened the regional concerns that price adjustments had always brought forth. When the freight increase went before the Board in 1948, seven of the nine provinces objected. Both national parties passed resolutions expressing concern, and the Government responded by appointing a royal commission on transportation (the Turgeon Commission).

Figure 26
CONSUMER PRICE INDEX, 1913-1951



Source: Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series K8.

The Turgeon Commission was the last one in which passenger issues remained marginal. As it noted, there were only a few isolated complaints about the quality of passenger service, largely from the island provinces, and one complaint about passenger fares from British Columbia that cited the mountain differential rate built into the standard fare. In this latter case, however, the railways accepted the idea of abolition and the differential rate soon disappeared.²⁵³

Despite the limited attention it gave to passenger transportation, the Commission is of importance here for two reasons. First, the Commission marked the return to the political arena of the hotly contested battles of the 1920s on regional equity and railways. Much of the same rhetoric was used and, if space allowed, one could show how provincial submissions returned to practically all the concerns of the earlier years.²⁵⁴ The origins of the Commission, which as it somewhat tartly pointed out were definitely political, demonstrate what one critic has termed the "ideology of freight rates."²⁵⁵ Second, the Commission is important because in its recommendations it supported a principle that in recent years had been used by the railways in their application for freight increases — cross subsidization. The sharp rise in expenses and the decline in passenger services were a growing concern of the railways immediately after 1945. At that time they did not seem to have had any intention of eliminating or even sharply reducing passenger services. What they did do, however, was argue that passenger service losses should be considered in freight rate increases. Freight rates, in other words, should subsidize the passenger service. The Commission agreed. "The freight and passenger services are essential and if the passenger fares cannot be raised to produce sufficient revenues to enable the passenger traffic to pay its own way the freight traffic must bear the burden."²⁵⁶

There are several points of interest in all this. First, the regional position on passenger transportation was negative: both the Western provinces and those in the Maritimes saw lower freight rates as more important than the railway subsidization of passenger service. This continued the long-standing practice of emphasizing freight while taking passenger service for granted. Second, the Turgeon Commission represents a break from the past in that the railways were now becoming concerned about their rising passenger deficits. That anxiety had prompted the railways to at least raise the question of subsidization of their passenger service. Third, all the parties involved seem caught between older and newer conceptions of the transportation

industry in Canada: the traditional view of railways as a monopoly to be controlled was challenged by a newer view that railways were a declining industry, faced with fierce competition and a shrinking percentage of the transportation market. Regulation designed to control a monopolistic industry was now turning to subsidy issues in the face of a troubled one. Amidst all this, as A. W. Currie pointed out, the Commission "displays the confusion of objective which has characterized the history of transportation in Canada almost from the beginning." Were railways businesses, entitled to maximize profit and pursue corporate goals, or were they agencies of social well-being in which profit was secondary?²⁵⁷

Through the 1950s more and more policy makers began to realize that the competitive position of the railway had indeed changed and that this was especially true in the case of passenger services. Previously, the public had not lobbied strongly on questions of passenger routes and the Board of Transport Commissioners as well as the politicians had generally accepted railway decisions on the services to be provided, increased or dropped. All that, however, had rested on the assumption that the overall level of passenger service was likely to remain the same or improve. As it became apparent that services were shrinking affected communities, the press and apparently the public at large became more insistent that the railway should be viewed as an essential service. Cases for reduction of service or abandonment proceeded through channels only slowly and with great outcry. This left the railways caught between irreconcilable forces. Passenger services were becoming more and more of a burden. By 1958 the railways estimated their annual losses on passenger service to be more than \$56 million.²⁵⁸ It was no longer feasible for railways to absorb such losses or for adjustment of freight rates to cover it. On the other hand, the slowness of the regulatory system prevented railways from adjusting services in an attempt to maximize efficiency. Uncertainty discouraged investment in new plant, and the problems of discontinuance made introduction of new passenger routes a very dubious proposition. The railways remained committed to passenger service, but for the first time in the history of the nation there were worries that railway passenger service might be unable to compete successfully in the intercity market.

This was clearly recognized by a royal commission on transportation established under the Diefenbaker government (MacPherson Commission). Though freight rates and their increase were the basic reason for the appointment

of the Commission, it, unlike its predecessor, found a great deal of public concern about the quality and future of passenger services. Numerous municipalities and interest groups complained that declining service, actual or planned, threatened their well-being. At the same time there was also a feeling among witnesses that passenger services were too costly and that something had to be done to rationalize the system, though usually at somebody else's expense.²⁵⁹ For their part the railways pushed hard to encourage a facilitated system of service abandonment and, where services "are held to be necessary in the public interest," to develop Government subsidy programs.²⁶⁰

In attempting to sort out all the concerns, the Commission started with a point not fully understood by the earlier Turgeon Commission. The problem, it concluded, was that both the Government and the public were behind the times in their conception of railways. Railways were no longer the giant monopolists of former years but faced instead a very competitive environment. Yet the regulatory system, the political arena and public pressure all worked against any flexible response by the railways. This was as true in passenger services as in freight.

In this [adjustment to competition] they were handicapped, not only by federal regulatory requirements, but also by public pressure which customarily took the form of intense resistance to the dislocations which might be occasioned by these adjustments. For example, such matters as the removal of passenger services operating at a loss or the abandonment of unprofitable branch lines proved virtually impossible to decide with reference to normal commercial operation.

At another point the Commission noted pointedly that the railways are accused of deserting communities by withdrawing passenger-train service, when a more objective view would be that the communities have deserted passenger train service.²⁶¹

If Canadian transportation were to adjust to the requirements of the last half of the century, the Commissioners concluded, the railways had to be given increased flexibility. They were no longer dominant in transportation and indeed, in passenger services, were now minor participants. If they were to continue to have any role in that area they must be left with the freedom to invest where they could make money and to get out of the business where

there was no money to be made. The whole thrust of the Commission was thus toward deregulation and, to the extent that regulations continued, a recognition that railways should no longer be singled out for regulation not faced by other carriers.

Yet, in a significant qualification, the Commission drew upon an idea that the railways themselves had argued. Canada had always faced the dilemma of requiring certain transportation services when the market itself did not justify such services. The response had traditionally been Government support in one form or another. Now that rail passenger services seemed to be in difficulty, it was only natural to apply the principle to that sector. This would be done in two steps. First, said the Commission, there should be a transitional period of four years in which the railways would be subsidized for moving slowly to rationalize services. Such subsidies would run from \$62 million in 1961 to \$12.4 million by 1965. By the end of that period, however, much of the passenger service in the nation was expected to disappear. Second, railways would thereafter remain on a more selected range of potentially profitable lines and would be subsidized indefinitely on those crucial passenger lines where "no reasonable alternative public highway exists."²⁶²

The MacPherson Commission is significant in the history of Canadian transportation, for it marked a bold departure from past thinking. First, it challenged many of the myths about the power of railways and the public complaints that railways must provide adequate passenger services. Railways were neither as powerful nor, given modern public habits, as essential as had been previously argued. Accordingly, the Commission accepted the railways' complaints about being forced to continue services that were no longer required or profitable. Second, in the name of the public interest, the Commission altered the nature of subsidies to passenger services. Cross subsidization from freight to passenger services would not be favoured in the future. Instead, the subsidies would be direct, with the Government accepting the responsibility and the taxpayer footing the bill in instances where services were thought essential. The combination of public subsidy and service reduction practically guaranteed that passenger services, which had so recently become politicized, would remain highly controversial in the future. At the same time the assumption was that railways would significantly rationalize passenger services. The up-front subsidy should ensure that the public and politicians did not get too carried away preserving non-essential passenger trains.

Things did not work out as neatly as the MacPherson Commission or its supporters had hoped. The minority governments that characterized the 1960s hindered the introduction of new legislation and discouraged bold steps. Amendments to the *Railway Act* were prepared under the Conservative government in response to the MacPherson Commission. In 1964, under a Liberal government, a version of this bill was actually introduced in the House of Commons, only to die in committee. By that time the new Minister of Transport, J. W. Pickersgill, took the logic of the MacPherson Commission one step further. If railways were no longer the dominant form of transportation in Canada, it made little sense to deal with the problem by looking only at railways. By 1966 the Department of Transport was ready with a complicated bill designed not only to deal with the specific problems of the railways but also to create an agency that could look at the issues of transportation from an integrated perspective.²⁶³ The result was the *National Transportation Act* (NTA), which became law in 1967.²⁶⁴

In principle the NTA adopted the concepts put forth by the MacPherson Commission as modified by subsequent public pressures and political concerns. Those modifications, however, proved to be significant. In terms of passenger transportation the Act extended the principle that the public interest had to be weighed (through hearings before the newly constituted Canadian Transport Commission [CTC]) before steps were taken to abandon branch lines or passenger services. Then, if service was deemed to be necessary despite losses, the government was obligated to pay 80 percent of such losses.

There were also important departures from recommendations of the MacPherson Commission. First, the Commission had not envisaged a "super-agency" to handle transportation matters. Second, the MacPherson Commission had clearly intended that Government support of passenger routes come into play only under relatively extreme circumstances, "where no alternative exists."²⁶⁵ The NTA took a more open-ended approach, however, and included such considerations as "future transportation needs" and the politically loaded possibility of appeal to the Governor General in Council.²⁶⁶ Related to this, the MacPherson Commission had not considered partial subsidies, but the drafters of the legislation were concerned that total subsidization would lead to profligate attitudes on the part of the companies. It was uncertain whether the Government, faced by contradictory arguments from the public and corporations, had decided the NTA subsidies were largely transitional to allow the rationalization and abandonment of passenger

services or part of a national commitment to low-cost, publicly supported rail transit. What was certain is that with the NTA the new reality of inter-modal competition and an ailing passenger rail sector had been recognized in law.

6. THE MODERN ERA: 1967-1990

The years after 1967 mark the most recent phase in the history of Canadian transportation and transportation policy. The many battles that have been fought over passenger transport in these years pose a danger for the historian. Never in the history of the issue have so many papers poured forth from the corporations, lobby groups and especially government. They range from careful and detailed studies of technical issues by the new Research Branch of the CTC to wide-sweeping generalizations about the importance of service by politicians protecting local constituencies. Yet from an historical perspective the period from the passage of the NTA until now is extremely brief. It is important to step back from the trees (or ex-trees now converted into memos) and try to single out the major landmarks and trends of the last 20 years.

From this more remote perspective, four things happened that set the stage for the various debates. First, within a very short time it became apparent that the Government was not going to apply the NTA subsidies stingily. The new CTC and the Government were cautious in their approach to railway abandonment.²⁶⁷ Second, the railways finally abandoned any hope that the passenger business could be revived. Third, the annual cost of the subsidies increased much more rapidly than had been expected. By the early to mid-1970s the Government was seeking to extricate itself from the subsidy system. Ministerial decisions in 1976, 1981 and 1989 to reduce passenger service reflected this Government desire to rationalize passenger rail and, not incidentally, to reduce the subsidy levels. Fourth, while the Government was searching for a solution to the ever more expensive rail passenger service, a new ambivalence was creeping into policy on roads and cars. Energy crises, pollution concerns and other issues meant that an improved highway system seemed a dubious alternative as a form of passenger transportation.

All of this activity led to considerable comment from a public that talked in one fashion and acted in another. People weren't leaving their cars to return to the rails, but they continued to see passenger rail service as desirable

and viable. There was a continued attachment to railway travel even among those who rarely or never took trains. The image of the railway corporation as the rapacious monopolist of yesteryear also created scepticism about railway motives in abandoning passenger service. Had they really tried to make it profitable? Was the bad service perhaps a deliberate response to subsidization? There was also confusion because the continued success of European railway systems and the increasing success of the Amtrak system in the United States seemed to contradict the Canadian experience of the last 20 years. Each attempt to cut back passenger rail service brought an angry response. It is thus in this modern era that the dichotomy between public attitudes and Government policy is most pronounced. It is in this era as well that there is a sharp gulf between the public consumption of passenger rail services and public support for the continuation of those services.

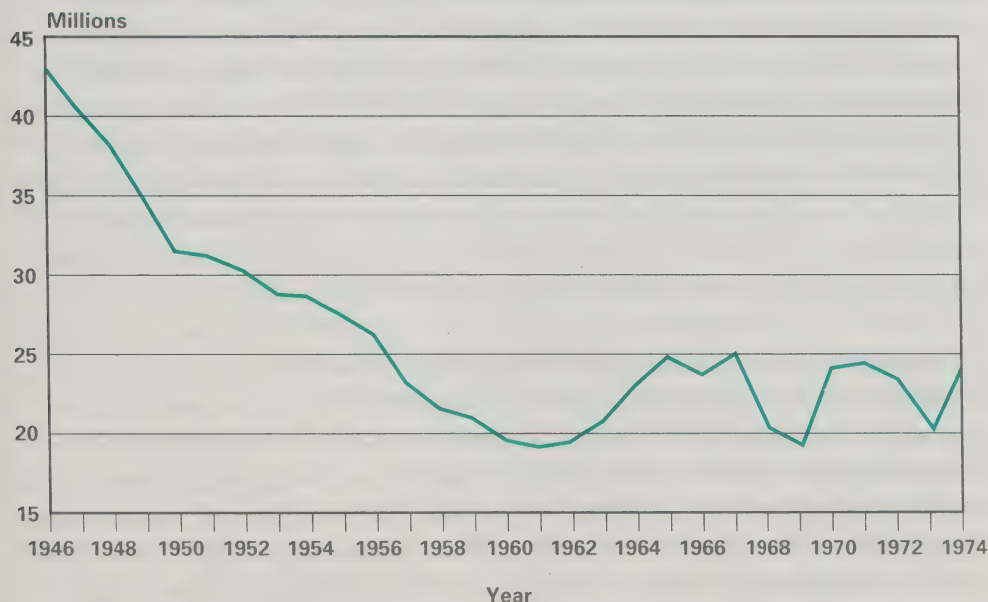
Another characteristic of the modern era has been the increasing influence and importance of other carriers. The airline industry went through another period of growth, and by the 1980s middle-income Canadians routinely used airlines for long-distance trips. For business it had become the standard mode of travel, even for relatively short distances. At the same time the automobile became ever more ubiquitous. Between 1961 and 1975 motor-vehicle registrations doubled, to one vehicle for every two Canadians. Theoretically, everybody in the nation could have been travelling in a car at the same time and have left the back seat empty! By the late 1970s all but the poorest of Canadian families owned cars.

Such patterns are apparent only in hindsight. In 1967, nothing was yet certain. Though there were accusations that the railways were deliberately chasing passengers away, every indication is that as of 1967 the railways were still attempting to make passenger service viable.²⁶⁸ Beginning in 1963 CNR undertook a number of initiatives, including a flexible system of pricing known as the "Red, White and Blue" fares. These were designed to encourage off-peak travel at lower cost. A variation of the idea was soon adopted by CPR. Both railways upgraded their transcontinental service during the decade. CNR also spent considerable money and effort trying to draw traffic on the heavily travelled Toronto–Montreal corridor. The introduction of the *Rapido* in 1965 made the route, at least according to the Company, the "fastest train passenger service in North America for a comparable distance."²⁶⁹ Plans for a new Turbo train called for a further reduction in time by the end of the decade. There were also inventive new partnerships

emerging as provincially subsidized commuter systems appeared in Ontario. The GO system, first appearing in 1967, marked the reappearance of an older role for the railway — carrying short-haul commuters. Commuters would form an ever-greater proportion of total train ridership through the 1970s.

There were also hopeful signs that the long decline of ridership was coming to an end. In fact, after a generation of decreases, the late 1960s saw an actual growth in the number of passengers carried (see Figure 27). A growing population, especially the large youth population of the 1960s, needed public transportation. From a low of fewer than 19 million riders in 1961, numbers increased to more than 24.5 million in centennial year. Over the next several years those numbers remained more or less stable, fluctuating between 20 and 25 million or, to put this in perspective, about one trip per year per Canadian.²⁷⁰ Finally, the NTA system of subsidies *seemed* to be working as the MacPherson Commission had intended despite the cautious approach taken toward abandonment of services. In the late 1960s and very early 1970s subsidies to railways declined consistently. By 1971 railway subsidies for passenger travel had been halved to \$53.9 million.

Figure 27
NUMBER OF RAILWAY PASSENGERS



Source: Statistics Canada, *Canada Year Book*, and railway company annual reports.

SUBSIDIES AND CUTBACKS

Yet all of this proved illusory. If there was ever a chance that passenger railway travel could be retained while the subsidy system was phased out, that chance disappeared with the changing economic fortunes of the 1970s. From World War II until the later 1960s Canada had experienced one of the greatest periods of sustained growth in its history. Any recessions were brief and not too severe. This meant that passenger transportation policies, whether in the public or private sector, were formed against the background of expectations of continued economic growth and prosperity. Then, in the early 1970s, the period of prosperity faded away. The 1970s were to be marked by growing "stagflation."²⁷¹ For the next decade economic growth remained unsteady and significant inflation created instability. One component of this stagflation was especially important to the transportation sector. The large-scale increase in energy prices in 1973 had ominous implications for transportation. Travel (or the carriage of goods) became more expensive relative to other activities, and Canadians curtailed travel or sought ways to make it more economical.

All sectors of the transportation industry were hurt by stagflation in general and by the energy-price shock in particular. The airline industry experienced some of its worst years ever in the early 1970s, and both the number of passengers and number of passenger-miles decreased — a very rare event in an industry known for growth.²⁷² North American car makers laid off thousands of workers. Consumers turned increasingly to smaller, cheaper and more gas-efficient cars. The airline and car industries, though they might suffer in the short term, operated from a position of relative strength. Their technology was still in demand. Car registrations did not decline, though people moved to smaller vehicles, and after the short retrenchment of the early 1970s, airline ridership again began to increase.

Long after the immediate impact had worn off, however, the energy crisis affected discussions of transportation in Canada. The concerns about conservation that the crisis awakened accentuated a public and official shift toward the automobile that was already under way. People would not give up their cars to any great degree, nor did the 1950s imagery of freedom and status disappear. However, a counter-image was now developing that emphasized the negative side of a society dependent on the automobile. A series of events, from safety concerns through energy and pollution concerns,

meant that the “absence of doubt” that characterized public investment in roads in earlier decades disappeared. In the last couple of decades there have been all sorts of doubt.

There had always been some critics of the suburban mythology, and the lamentable safety record of cars had been a cause of concern in the 1950s as much as in other decades. In those years, however, the reply to the critics was not to fetter the car but to design it or its transportation routes better. Modern highways would provide greater safety and better separation from local streets and generally improve the quality of life. Underlying these attitudes towards progress through to the mid-1960s was the belief that technology was a positive force. The unleashing of corporate and governmental inventiveness would, as was discussed earlier, improve peoples’ lives. By the late 1960s this was an increasingly challenged proposition. For the age was, in many ways, dedicated to a cynicism about material progress and to the developmental orientation that had gone before. The North American automobile, as the symbol of the modernism and technology, was bound to come under criticism.

The critique came in stages. Initially, the car lost some of its personal glory as a symbol of technology’s benefits. The North American automobile, so glorified by an earlier generation, turned out to be a shoddy example of technology. The publication of Ralph Nader’s *Unsafe at Any Speed* in 1965 helped create a new ambivalence toward the North American car.²⁷³ Technological progress, said this new perspective, was largely illusory — a matter of design and gimmicks. Underlying it all was a cavalier attitude toward such basics as safety and reliability. People still loved their cars but they were no longer certain they could trust them.

A second source of concern arose because the pressure for highway development was moving to the cities. Municipal expenditures on streets continued to climb in the late 1960s even as those by federal and provincial authorities levelled off or declined. Yet the car posed real problems to urban life. The vision of gargantuan expressways bisecting city centres was increasingly unacceptable. From the politicians’ point of view it was too expensive. At the unofficial level, citizen resistance grew as city neighbourhoods feared they would be sacrificed in the name of suburban traffic flow. The highly publicized fight to stop the Spadina Expressway in Toronto in the early 1970s and the debate over the “Third Crossing” of Burrard Inlet in

Vancouver made it apparent just how politically charged such issues could become.²⁷⁴ Thus the combination of urban problems and contemporary sensibilities made the automobile-oriented transportation system seem increasingly dubious, at least in and around the big cities, by the early 1970s.

Neither Nader's crusades nor the urban concerns of the early 1970s, however, had as much impact as the energy crisis of 1973. Then critics of the car could claim not just that the vehicles were badly made or that urban congestion was a problem but that the very act of driving a car was socially destructive. Numerous popular books proclaimed that the energy crisis had brought the North American automotive dream to a crashing halt. One work, attacking the automobile society upon which North American development had been based for the last quarter century, even gained the imprimatur of the Governor General in the form of a foreword. It was a sign of the times.²⁷⁵ So too was the new image of the car. From the symbol of freedom it had become the metaphor for rapaciousness and inefficiency. "I should like to dedicate this essay," said one critic in a book aimed at engineers, "to the majority of mankind which has no hope at all — and, I sincerely hope, no ambition, of owning an average American car. This paragon of ecological virtue has an astonishing capacity for swallowing non-renewable resources. Quite apart from its contribution to the death rate, air pollution, noise and blighting of the city and countryside, the typical U.S. car spends much of its time transporting 1.5 people."²⁷⁶

Criticism also moved inside government circles. The official Canadian position became, at least implicitly, anti-automobile. Energy conservation was, for at least a decade, a matter of official concern and exhortation to the public. An Office of Energy Conservation was established, and the Canadian government threw its support behind measures to conserve energy. The automobile was, as everyone knew, something less than energy efficient and therefore governmental efforts at energy conservation meant government efforts to downgrade the place of the car in Canadian transportation. Federal departments like Energy, Mines and Resources warned that Canadians must abandon their dependence on the car and move "from single passenger cars to multi-passenger cars, buses and trains, and greater use of public transit systems."²⁷⁷ A Department of Transport study referred to the car as "the most dangerous mode [of transportation]; a major source of congestion in high density areas; inefficiently used in terms of loads and peak periods; the largest source of pollution and energy consumption;

and most subject to weather vagaries.”²⁷⁸ The Science Council of Canada wrote a series of reports through the decade that urged the preservation of alternative forms of transportation and warned of the impending crisis in energy supplies.²⁷⁹

We could go on with examples but the point is fairly obvious. In Canada as in the United States the basic perspective on the car had changed.²⁸⁰ The critics, previously in a minority, had now become the prevailing wisdom. That wisdom could be summed up to the effect that even if the car was not inherently bad as a transportation medium (and that was a point of debate) the way the North American car culture had evolved was socially and economically destructive. Overly large cars — often used by single passengers — vast freeway systems and excessive commuting distances had created problems of safety, urban pollution, energy depletion and social dispersal. Some of these issues could be ameliorated by better and more efficient cars, but the preferred solution was alternative transportation systems. Certainly none of the problems could be solved by building more freeways. That only encouraged more cars.

The result was the end of the single-minded focus on cars and roads that had characterized the past decades. In Ontario, for example, the Department of Highways was, in 1971, merged into the much larger and diverse Department of Transportation and Communications. “Transportation today, however, is not confined to highways. It has acquired an all-encompassing role,” concluded the final report of the old department.²⁸¹ Two events were especially symbolic, however. The first involved the Canadian Good Roads Association, which dates back to 1914 and, through the Ontario Good Roads Association, to 1894. Over the years it had been the centre of discussion on and support for improved Canadian roads. Yet in 1970, arguing that “we in CGRA are aware that roads alone will not solve our transportation problems,” it changed its name to the Roads and Transportation Association of Canada.²⁸² Second, Statistics Canada had, since the late 1920s, published figures on road expenditures in Canada. The beginning of such regular information had marked the new rise to importance of the road in Canadian life. In 1976, however, it discontinued the series. It was a symbolic reflection of the decline in the perceived national importance of road construction.²⁸³

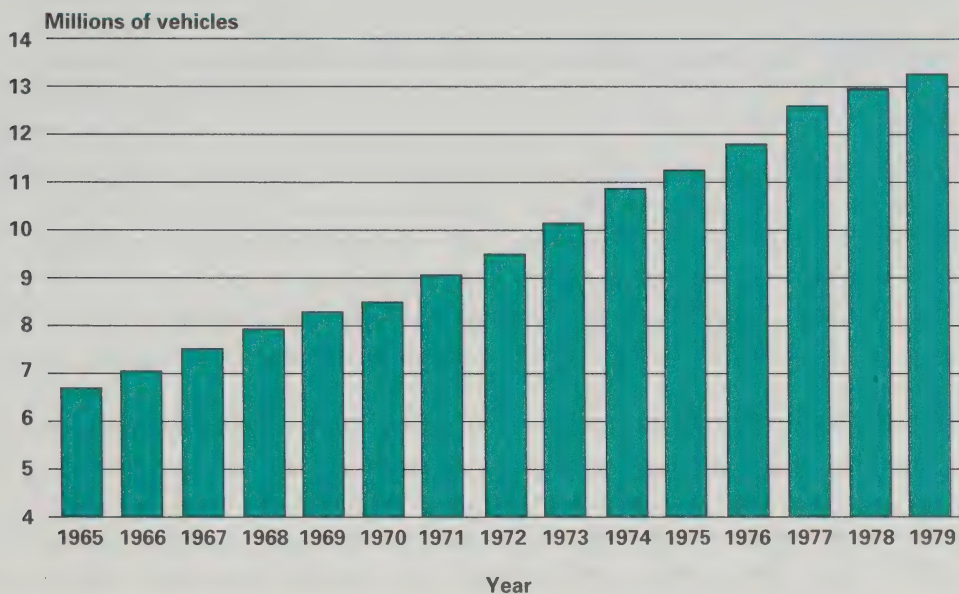
The shift, though important, must be qualified, however. The 1970s saw paradoxical pulls exerted on governments. Despite the concern with energy conservation and despite the talk of alternative modes of transportation, the

number of cars continued to increase in Canada (see Figure 28). The public was not going to be wooed away from the car, and thus governments were torn between searching for alternative forms of traffic and responding to ongoing congestion on the roads. This made transportation policy difficult to define. The prevailing wisdom was to woo the people out of their cars. In some places efforts were made, with some success, to do so. The GO train system in Ontario, for example, provides a perfect example of the shifting priorities of the provincial government.²⁸⁴ GO was an exceptional system, however, based in the largest metropolitan concentration in Canada. Most places had neither the provincial revenue nor the population density to support such a system.

The rise of concern about the automobile was paralleled by declining investment in the road infrastructure of the nation. As late as 1966 *Road and Wheel*, the magazine of the CGRA, proclaimed that Canada was headed for a record year in road building, characterized "by large and more complex projects." The next year the headlines trumpeted that construction had reached \$1.6 billion and that the record pace was continuing.²⁸⁵ In fact, however, changes were in the air. In 1967 both federal and provincial road expenditures decreased in real terms. The provincial cut was especially severe: it was the greatest percentage drop since the great boom had begun at the end of the war. The next few years saw the trend continue. In real dollars, expenditures on rural roads would not reach 1966 levels again for more than a decade (see Figure 29). Even then the recovery was short lived, and by the 1980s real-dollar expenditures had plunged once again.

The near-disappearance of the federal government from road construction was an especially dramatic aspect of the turnaround in road policy. Once the Trans-Canada and Atlantic Development grants wound up by the 1970s the federal government was little involved in highway policy. Expenditures fell in absolute terms to only a small portion of what they had been a decade before. In relation to provincial highway expenditures they became insignificant, declining from their peak of 16 percent in the mid-60s to less than 5 percent by the early 1980s and to less than 4 percent by the end of the decade.²⁸⁶ The basis of involvement had also become less unified, appearing for the most part under the Regional Development program under a series of ad hoc non-legislated arrangements with provinces known as ERDAs (Economic and Regional Development Agreements). There was nothing in size or cohesion to compare with the Trans-Canada program or even the Roads to Resources program of earlier decades.

Figure 28
 AUTOMOBILE REGISTRATIONS, 1965-1979

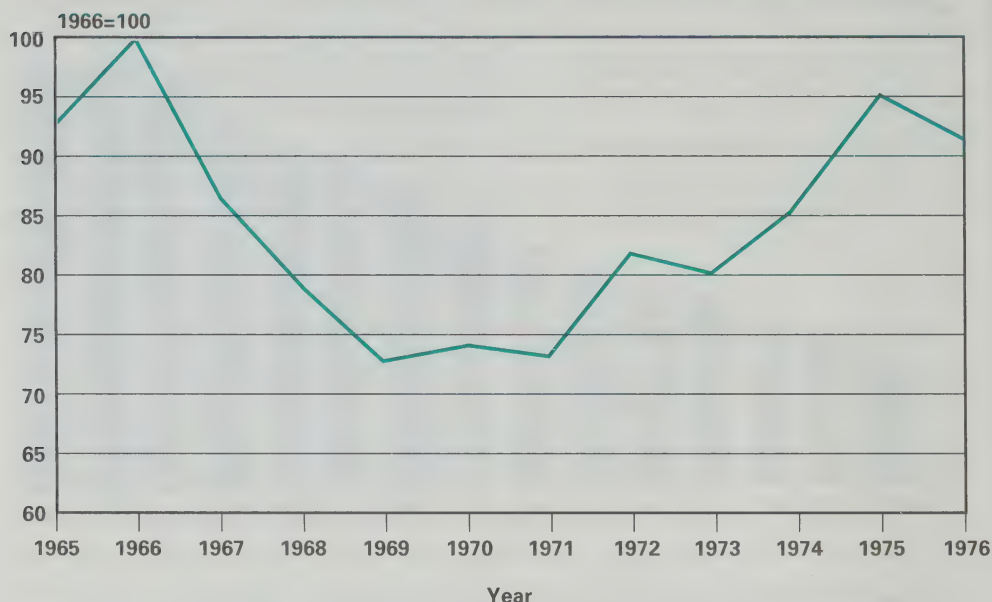


Source: Derived from Statistics Canada, *Vehicle Registrations*, Catalogue No. 53-219.

Growing doubts about the car culture complicated the debate over the decline of rail passenger service. As issues of energy conservation and pollution came to the fore the train was looked upon relatively favourably, especially when compared with the automobile. The railways were obviously unable to continue without large-scale subsidization and yet, as Via 2000 and other pressure groups reminded the Government, the railway was exactly the sort of alternative that the nation needed if the car was to be used more sparingly. Much of the argument in favour of the train is explicitly or implicitly based on the “social costs” of the car and hence the desirability of Government support for national rail passenger service. By the mid-seventies federal subsidies of passenger rail service were generally higher than federal contributions to highway construction. This was a reversal of orientation from the 1960s. Then the federal involvement in the Trans-Canada system and the MacPherson Report’s recommendation that the federal government get out of the railway subsidization business seemed to point the way to the future.²⁸⁷

Figure 29

**THE END OF THE HIGHWAY BOOM: PROVINCIAL PLUS FEDERAL
REAL EXPENDITURES ON ROADS, 1966-1976**



Source: Statistics Canada, *Canada Year Book*, 1966-1978.

Non-economic or non-direct economic arguments could be made that the railway should be preserved not in the name of nostalgia but for the social benefit it could confer.²⁸⁸ Despite this, the economic instability of the 1970s was ultimately disastrous for the rail passenger industry. The decrease in travel, along with the more competitive alternatives to rail passenger service, meant that the railways' hopes of finding a core of profitable routes proved illusory. As Otto Lang, Minister of Transport in the mid-1970s, put it, "there were, in fact, no profitable routes."²⁸⁹

The absence of any profitable routes made unworkable the principles underlying both the NTA subsidy system and the CTC caution toward reduction in services. The hope had been that profitable routes could, with the assistance of conservative rationalization and subsidization, support the non-profitable routes. Now it turned out there was no cross subsidy. The railway companies were left in an economic "no-man's land," directed to maintain a high level of railway services, subsidized for 80 percent of their losses but

required to pick up the difference. Moreover, even if the subsidies had been total there was still an opportunity cost. Every railway prediction was that freight traffic was going to increase continuously over the next several years. Every dollar, every locomotive and especially every mainline track that was cluttered up with money-losing passenger services prevented railways from meeting the demand on the freight side.²⁹⁰ The absence of any profitable routes was equally disturbing to Government officials. The basic problem can be seen in the subsidy figures. After reaching the low point of \$53.9 million in 1971 Government grants to railways began to go up. By 1973 the subsidies had increased to more than \$130 million; by 1974, to \$164 million. Also, an unprecedented two thirds of these subsidies were now going to passenger services. This was not something envisaged by either the MacPherson Commission or the framers of the NTA.²⁹¹

These forces set off a chain reaction. Until now at least CNR if not CPR had maintained an interest in passenger services. By the mid-1970s, however, the railways spent most of their time trying to get out of the business. In the Government there was a growing cynicism about the future of passenger services. It was a technocratic era, and the romance of railway travel seemed to have little impact upon the Department of Transport, which looked increasingly to the bottom line cost of maintaining services for a shrinking clientele. "Described in extreme terms," said one mid-1970s study from the Department, "it would be cheaper for the government to buy a bus ticket for each rail-passenger and give it to him free of charge, than to subsidize the rail-passenger traveller as at present; and bus service is available alongside most of the rail routes involved and in the case of many routes, air service as well."²⁹²

Until then political leaders had been very cautious about any policy of rationalization. Jack Pickersgill had captured the flavour of the political view of railway matters when he commented in 1966 that "transportation has the capacity to arouse emotions and prejudices . . . more completely than questions of race, religion or language."²⁹³ It is not surprising that his successors had been reluctant to pursue a policy of passenger rail reductions. The CTC, taking its cue from the political and public mood, had also been cautious in its approach to reductions in passenger service and branch lines.²⁹⁴ The rising deficits and the studies within the civil service had their impact, however. By 1976 the politicians were convinced. In January of that year the Minister of Transport, Otto Lang, announced a transport policy aimed at

rationalization of railway passenger service.²⁹⁵ This decision, to accept the inevitable public criticism and to try to reduce the cost of railway services, is important. What Lang did was to force action on an agenda that had been in the making since the MacPherson Commission. Passenger services would be reduced and the political consequences would be accepted. A core of services, however, would be preserved and upgraded. Ever since this statement was made debate on transportation policy in Canada has revolved around the wisdom or folly of the actions put in motion that day.

Formally, Lang did two things through the policies he enunciated in January 1976. First, he instructed the CTC to look at passenger rail service with an eye to rationalization.²⁹⁶ These instructions gave three basic roles for passenger trains: as short- to medium-distance carriers between cities; as commuting trains; and as transcontinental carriers providing a high level of comfort.²⁹⁷ In effect, this put the emphasis on the Windsor–Quebec City corridor and on the transcontinental trains. This left the CTC with a difficult task. It did not know specifically what levels of service or subsidy the Government considered reasonable, only that rationalization was desirable. Nevertheless, the CTC responded to Lang's directive by proposing reductions in service and holding public hearings across the country. After considerable controversy and much press coverage, the CTC proposed a series of reductions, mainly affecting underutilized regional routes.²⁹⁸

The second thing Lang did was remove the burden of passenger service from the existing railway corporations. He did this by using the unusual method of an appropriation vote to create VIA Rail in 1977. In 1978 it became a Crown corporation.²⁹⁹ If the Government was to revive passenger service given the disinterest of the railway corporations, such an agency made sense. It ended the pretence that railway passenger service was still a private business. The Crown corporation would, presumably, be more interested in providing adequate passenger service for the simple reason that that was its purpose. VIA was not without its critics, however. Then and since, defenders of rail passenger service have argued that VIA was without a legislative mandate, was unsupported by the Government and was thereby destined to fail to attract passengers.

The cuts and the creation of VIA were met by new counter-balancing forces. In Ottawa, and then across the country, a group of individuals drew upon a British precedent to create a lobby group in support of alternative

transportation. The existence of Transport 2000 and other lobby groups ensured that in the future Government moves on rail service would face close scrutiny. The 1977 reductions in service also had an impact upon the public at large. As we have seen, the public hearings and the resultant newspaper coverage awakened public interest.³⁰⁰ Certainly the hearings in 1977 generated a great deal of attention and controversy. This was especially true in the Maritimes. Both the transport needs of the region and the tradition that tied railway service to the Confederation pact made for lively discussions when the CTC Rail Committee came to the region in May 1977 to discuss rationalization.³⁰¹

There were also forces in the public sector that began to reconsider rail service. If nothing else, there was VIA Rail. Its very purpose revolved around the provision of passenger service. Its existence, therefore, created an additional voice in support of adequate passenger service. Immediately upon organization VIA undertook a large-scale marketing campaign that succeeded in increasing ridership and operating costs dramatically in the first couple of years of operation. This emphasis not on profit but on service and ridership is reflected in the Corporation's annual reports of the first decade. These reports consistently point to the social and political significance of passenger transport. They also show that VIA is aware that its audience is the public service and the politicians rather than the shareholders of a private corporation. "Now here are we," wrote the President in the 1985 *Annual Report*, "the present inheritors of that remarkable tradition, feeling somewhat insecure or uncertain about whether the long progression is still relevant." The service still had a place however. "One can only conclude that as a nation we have not yet found the right way to think about passenger rail. Yet find it we must, unless Canada is to lose a vital link between its citizens. . . . Surely we should not break a tradition that is emotionally important to many Canadians because of its pivotal role in our history."³⁰²

In the meantime the public service tried to grapple with the issue. Rail passenger service was costing the public treasury a great deal of money, but the automobile's dominance continued to raise social and energy concerns. Indeed, the striking thing about Government studies in the later 1970s and early 1980s is the amount of time spent on the subject of that most anaemic of all modes, the passenger rail system.³⁰³ In effect the controversy of 1976 and 1977 had raised in a serious way the question of the mix of transport

modes in the Canadian future and had done so when the car was under greater criticism than ever before as a wasteful, dangerous and expensive form of mobility. In the latter part of the 1970s, reported the Science Council of Canada, the car "captured approximately two-thirds of the annual growth [of the transportation market], with the other third going to air travel. At a time of energy conservation this is not happy news."³⁰⁴

In the past decade the other major force driving Government policy has been the national deficit. Through the 1970s the deficit increased dramatically. By the early 1980s it was becoming a significant, negative economic force and a political issue. Simply put, cynicism about the future of rail travel drove Government policy in the 1970s. In the 1980s the despair lessened as lobby groups, VIA Rail and others began to look for ways to rejuvenate the system. Even as the fatalism toward passenger rail lessened, however, the imperatives of Government finance took over. The 1980s were not times conducive to massive and unproven investments in passenger rail technology. Instead, Government watched as total subsidies to VIA Rail increased to more than \$300 million by 1980 and to more than \$600 million by 1982. Attempts to reduce the deficits in 1981 met with massive public reaction, more hearings and polls that showed that the majority of Canadians wanted rail service maintained.³⁰⁵ The Opposition Party commissioned its own task force and created a lively defence of railways that received considerable press and public attention.³⁰⁶ The overall pressure was so great that in an equally controversial move many of the reductions in service were subsequently restored.

To some degree Government policy was floundering by the early 1980s, being pushed to and fro by contrary pressures. Public opinion ran against reductions in service, and there was no consensus among the government's advisors on direction that passenger transportation should take. Yet the rising national deficit and the rising railway subsidies prevented any sort of bold initiative. VIA Rail was widely seen, justly or not, as half a solution: a company without a mandate and without adequate funds. An Ontario task force report was especially critical, arguing that "VIA has no effective ability to operate passenger rail services in Canada, little control over systems and costs, cannot set its own budgetary priorities or effectively plan or experiment with different levels of service."³⁰⁷ VIA was now by far the most expensive part of the railway system for the taxpayer, absorbing nearly three quarters of all subsidies.³⁰⁸

Against this backdrop of uncertainty, the most recent effort to break the logjam came with the newly elected Government's rapid move to alter the whole approach to passenger transportation in Canada. Feeling that the complex set of regulations and hearings inhibited innovative solutions, the Conservative government moved toward a simplified regulatory system where necessary and a deregulated system where possible. The idea was for "less regulation, leading to less government interference, [which] will encourage innovation and enterprise."³⁰⁹ This policy, which was reflected in the new *National Transportation Act, 1987*, is still evolving and is therefore beyond any historical study. The new approach had effects on passenger transportation, but to uneven degrees. Rail service was not really affected that much. VIA, after all, had as its primary role the provision of passenger service rather than profit. It was cost conscious to a degree and for some time had been emphasizing "corridor services" to increase revenue. Deregulation did not change that emphasis and did not open the way to wholesale line abandonment. So long as the subsidies were maintained there was no purpose in doing so. In fact, when the next and most recent round of rail service cuts came, they did so in the traditional way and for the traditional reasons. The Government ordered the cuts made because of the increased deficit.³¹⁰ Likewise, the process of cuts engendered the same round of studies, controversy and outrage by opposition parties that had met both the 1977 and 1981 cuts.³¹¹ All of this indicated, of course, that passenger rail service was far beyond the state where it could be considered in normal free-market terms.

There was a second effect upon passenger service, however. Deregulation had a significant impact on the airline industry. Despite the temporary setbacks in the wake of the 1973 energy crisis, the industry had seen healthy growth. Between 1972 and the mid-1980s passenger figures doubled, reaching 29 million by 1986, far eclipsing the number of train passengers.³¹² The rapid shift that took place is brought into perspective by recalling that only 20 years before, five times as many people travelled by train as by plane. Thirty years before, 15 times as many did. Airline traffic continued to expand for many reasons. The standard of living had increased by the mid-1960s to the point where a larger percentage of the population could afford air travel. The improved technology and possibilities of wide-bodied jets allowed better service, as did fuel-efficient jets. Nor was flying the unknown quantity that it had been 20 years before. Most people took it for granted and used the technology freely. Increased volumes, in turn, allowed the

appearance of regional airlines and smaller planes that aimed to draw smaller communities into the air network and thereby displace alternative modes of travel.

Perhaps the most important factor increasing air travel was the effective lowering of costs to the consumer with the introduction of discount fares. In the early 1970s, as part of "package vacations," the CTC allowed airlines to introduce "charter-type fares" on scheduled services in 1977. These fares, hemmed around with restrictions on booking, cancellation and time of travel, nevertheless attracted economy-conscious travellers to airlines. Then, beginning in 1979, seat sales were introduced.³¹³ The initial conditions became successively less onerous, especially under the trend to deregulation in recent years. Thus, even before the mid-1980s move to widespread deregulation, the CTC had given the airlines much more flexibility. As a result, rail travel and to a degree even long-distance bus travel lost much of the cost advantage they had always possessed.

When sweeping deregulation was introduced in the mid-1980s, the airline industry was by far the most affected. Accordingly, there was a great deal of scrambling as the two major carriers (Air Canada and Canadian Pacific) sought to increase their market share. There were also dramatic changes in the field as the successful charter company Wardair entered and then failed in the scheduled market. There was also the purchase of Canadian Pacific by the aggressive regional carrier Pacific Western (and later the purchase of Wardair) to form Canadian Airlines International. As well there were the rise, fall and reorganization of several regional carriers. All of this is far too recent to assess in any definitive way. Two things are apparent, however. First, the highly competitive struggle for market share brought at least a temporary lowering of real air travel costs through much of the decade, further hurting VIA's ability to compete. Second, the oligopolistic situation of the airline industry and the regulations that carefully divided up routes had provided the industry with financial security and stability for the first 40 years after the war. Obviously this is no longer the case. Deregulation, the globalization of the industry, the projected open-skies agreement and congested airports have turned stability into instability. Speculation abounds that there will be a failure of a major airline, leaving Canada with only one national carrier within a decade. It is all very reminiscent of the way in which CPR and CNR were viewed in the interwar years.³¹⁴ Certainly, no

transportation system can be set up for the future that blithely assumes airlines will be able to bear the national burden any more than the railways were a generation before.

7. CONCLUSION: CONCERNS AND MYTHS IN CONTEMPORARY CANADA

As Canada moves into the 1990s public transportation is in a state of flux. The crisis of the railways has not been resolved, and there is still a lively debate between those who see rail services as essential to our future and those who see their abolition, at least outside the corridor, as essential to national solvency. The public seems to want better rail services but is increasingly pessimistic that this is a realistic expectation. The airline industry has probably never been so unstable as in the past few years. Corporate takeovers, the whole rise of the hub and spoke system, the controversy over the Pickering airport a few years ago and the growing congestion of places like Pearson Airport hint at limits to growth in the immediate future. Thus, though the technology of air travel will continue to be essential, it is uncertain what forms of corporate or route reorganizations will take place as the industry adapts to new circumstances. Buses continue to absorb a significant degree of travel, some of which would formerly have been undertaken by the train. Still, their services remain largely unintegrated with other modes, and over long distances they remain favoured only by those in lower economic groups. Only the automobile continues with relentless force to expand its dominance over the lives of Canadians.

This instability has led to a search for new policies, yet here too contradictory forces buffet decision makers. Lobby groups, such as Transport 2000, have continued to hammer home an effective defence of passenger rail and have been supported by other lobby groups, by opposition parties, and by other levels of government. In the meantime the Transportation Association of Canada has argued that the federal government must return to large-scale road investment if Canada's infrastructure is to be maintained.³¹⁵ The public is in the most paradoxical position of all. For it is the public's choice of transportation modes, especially its overwhelming attachment to the car, that has determined much of the policy of these years. The public is also one of the staunchest defenders of passenger rail transport. In 1981, when a series of VIA cuts were introduced, a Gallup poll indicated that a plurality of Canadians in *every* region wanted service retained or improved. Even

when people were asked if they would be willing to have taxes increased to support passenger rail service, the support continued in every region except Quebec.³¹⁶ The situation had not changed eight years later. When the Government moved to reduce VIA services in 1989 a poll indicated that an overwhelming 89 percent wanted VIA continued; this time there was majority support in every part of Canada. A clear majority (54 percent) actually wanted VIA service expanded.³¹⁷

Such support indicates that Canadians, even those who don't use the train, have a strong set of emotional attachments to passenger rail traffic. To understand why such attachments exist it is necessary to move from rational cost-benefit analysis or specific patterns of use to the realm of myth. A myth in historical terminology is not necessarily false, though it can be. Rather, a myth is a collection of beliefs and assumptions held to be true by a wide segment of the population, whether true or not. Canadian railways, because of their long history and recent controversial debates, are associated with a number of popular myths. Individual myths or elements of myths appeal to different segments of the populace and are used by different interest groups. As a totality, however, they possess a powerful collective hold over the thinking of the Canadian public and explain why policies to cut back services have been resisted so strongly.

The first and most potent of the myths is the nationalist myth. It developed around the time the Canadian Pacific was constructed. It was strengthened over the years by a combination of railway image building, public association of railway travel with "seeing the country," academic assessments of the importance of East-West links, and popular portrayals of the romance and adventure of the railway. A Gallup pollster summed up the power of this myth in response to the overwhelming support of VIA in 1989: "The trains are almost folklore in Canadian history, going back to the opening of the West and the national policy of John A. Macdonald."³¹⁸

There are two striking things about this myth. First, though most of the railway's business and its most important national role have always been freight, it is passenger service that evokes the emotional response. It is not enough for wheat to travel East to West — people must. Second, it is striking how populist this myth is. Many of the arguments about railways are technical and are, at least in detail, for the experts and enthusiasts. The nationalist myth is something that is not determined by expertise or by the special

interest groups, though of course it has been encouraged by them. It simply reappears at every point in which the public has input. The most recent example comes from the Spicer Commission's "hot line" and citizen forums. People resented the dismantling of such "national symbols" as the Canadian Broadcasting Corporation and passenger rail.³¹⁹

Also, though all rail passenger services are lumped together under the national myth, at the real heart of that myth are the transcontinental trains. "They are," concluded the President of VIA, "a unique and historic national symbol, a means of enhancing national unity and a contributor to tourism."³²⁰ "The closing of trans-Canada passenger service on the Canadian Pacific Railway," argued Transport 2000 recently, "is both a real barrier to communication and a symbolic attack on national unity."³²¹ "VIA Rail — National dream fades," trumpeted the headline in the *Toronto Star* during the most recent round of cuts.³²² Thus, public mythology can be focussed. The attachment to trains is really an attachment to passenger services and in particular to the grand tradition of the Canadian transcontinentals. Moreover, this nationalist attachment is not predicated on the belief that a rail service will save the country. Rather, it is seen in conjunction with a number of other forces as a part of the fabric of Canada.

Subsets of this national mythology exist in each of the regions. Sometimes it is purely a localized version of the national romance of railways. The 1977 cuts led journalist Jean-Paul Desbiens to reminisce about trips from Metabetchouan to Quebec City. Significantly, he saw his personal experience as part of the national experience, saying that it was "Trains: to people and to the land, a heritage. He remembers them well."³²³ In most cases, as we have seen, the regional variations emphasize the relationship of that particular region and the "national compact" — a compact that includes train services. Behind this is a more general principle that in the regions emphasizes the national policy role of the railways, or what the MacPherson Commission referred to as the tradition of obligation.³²⁴ Emphasizing this aspect of the railways subordinates profit or economic rationalization in favour of regional equity and national policy. As well-known Manitoba historian W. L. Morton testified before the MacPherson Commission, regulation of railways "was necessary to ensure that railways indeed fulfilled their role in national policy by providing the various regions of Canada with railway services on terms and rates equitable as between the public and the railways and as between the different regions of the country." An economist from

Saskatchewan saw the national myth by definition as uneconomic and therefore argued that railways should not be thought of in purely economic terms. "It is from the attempt to reorient trade from its economic pattern that most of our subsequent difficulties have arisen."³²⁵

A second and quite different myth exists about rail passenger service. This, for lack of a better term, might be referred to as the "if only" myth. The argument here is that in fact rail passenger service is viable if only some certain steps are taken or some previous arrangement had not been made. If only tracks were upgraded to allow faster service. If only car users had to pay their fair share of road costs. If only the railways had not charged VIA such exorbitant rates for running rights, old rolling stock and so on.³²⁶ The list could go on, but the basic point can be made in one citation. "We agree with Transport Canada that current rail service is untenable. The erosion of passenger services is real but given the virtual neglect of the rail mode over the past twenty years, it was inevitable."³²⁷ In one form or another all the "if only" myths rest on the belief that the decline in rail service came from neglect, bad planning or hostile, air-minded bureaucrats.³²⁸ Also essential is the corollary belief that good planning, effort, investment and a far-sighted government can turn the decline around.

The interesting thing about this myth is that unlike the previous one it is testable in normal economic and planning terms. Cost-benefit analyses can be done on technology improvements, alternative routings, integrated stations, an altered VIA mandate and so on. What makes this a myth and not just a proposition to be tested is its endlessly varied possibilities. Either side in the debate can easily maintain a position for or against rail travel whatever proposition is proven or disproven. Other issues are simply introduced. As a parliamentary committee reported in exasperation during the 1989 debate on cutbacks, "the Committee was confronted with a plethora of divergent statistics put forward by various witnesses with great persuasion and conviction. . . . The observation the Committee would like to make is that statistics can be used to make anyone's case or position sound persuasive."³²⁹ The statistics are just a subset of the whole "if only" debate. What is important here is not the truth or falseness of any particular assertion (though ultimately it is very important) but the very persistence of the debate. It reflects a widely held belief that neither the railways (for reasons of policy) nor the Government (for reasons of cost) has given passenger service an adequate opportunity to prove itself.

The third myth could be described as the social-benefits myth, and its characteristics borrow from both of the other major myths. The argument here is that whether or not passenger rail travel could be made profitable, it is worth the subsidy for social reasons. The 1970s argument about passenger service in an energy-conscious community that we have already discussed is a primary example of the myth of social benefits. More recently energy has become less of an issue, but another environmental issue, pollution, has replaced it as a focal point of discussion. Such environmental issues are joined by a myriad of others, some more reasonable in their claims than others. Thus, for example, train service is often seen as a means of transportation equity: the poor, elderly, and those with disabilities or from isolated communities are more likely to need the train than others. Regional equity is also involved, but more important recently has been a hinterland-metropolis approach in a more general sense. The marginal communities, whether in small-town Ontario, the Maritimes, or the North, are seen as dependent on trains in a way that large urban centres with good airports and roads are not. A 1977 editorial in the *Toronto Globe and Mail*, "Count the Social Costs Too," which condemned reductions of train service into Sudbury, perfectly illustrated this very potent concern. Cost efficiency, it said, left the Northern communities feeling "abandoned, neglected." Subsidies were especially necessary to prevent "inconvenience to thousands of small communities." Then, in an appeal to the national myth, it concluded, "We would never have built the railways if Otto Lang had always been in Transport."³³⁰

As was just mentioned, the social-benefits myth touches upon both the national identity myth and the "if only" myth. Canadians are well attuned to the general themes of pollution, energy consciousness and regional equity. To the extent that these themes are hammered home by supporters of passenger rail, therefore, trains become a "good thing" deserving support. This makes the national symbolism of the train all the stronger. Not only is it a part of our history, but as an institution, it still deserves support from the public purse. Likewise with the "if only" myth, the belief is reinforced that if it were only possible to measure and charge the true social and individual costs of such modes as the car, passenger rail would thrive, perhaps even without subsidy!

There are lesser social-benefits myths as well. An element of train travel is the sense of community it brings. Neither the automobile, with its privacy, nor the airplane, where people largely stay in their seats, gives the sense of

interaction that the train provides. "You meet such friendly people," recounted a New York couple in Jasper recently. "We almost hated to get off the train to come here [Jasper Park Lodge]." ³³¹ This sense of interaction and community appeals, at least in theory, to a good many Canadians, even when they do not take the train with any regularity.

Finally, there is what might be termed an "if . . . then" myth, which reflects a fear of what the decline in train service means to the community involved. It is not that the loss of a passenger train or so would actually seriously undermine the community but that the loss of passenger service is taken as part of a wider economic decline. For this reason it is the smaller cities, often in the Maritimes, that fight the hardest to retain a service. Their people do not use the service more than anyone else, but the jobs the railway provides, even if relatively few, and the symbolism involved in shutting yet another link between this community and "mainstream Canada" make railway closures an emotional issue. This theme applies equally to air services. Adequate train and air services to major points symbolize a city's status, are seen as essential to encourage growth, and are a part of the employment base of the community. It is the one area, therefore, where there may be a direct trade-off between trains and planes. Better plane service could be exchanged for reduced train service. ³³²

It is beyond the scope of this paper to unravel or test the many complex beliefs contained within these three major myths. Indeed, many of them are untestable, and many others, such as the national symbolism of the railway, do not lend themselves to specific cost judgements. The point is that they exist. Some may be accurate, some not. They reflect a strongly consistent part of Canadian public psychology, however, and may explain why a transportation system such as rail passenger service, so long underfunded and underutilized, has refused to die. It may also indicate that any blend of transportation systems developed for the future will have to take into account what the public expects of its national transportation system in psychological and social as well as economic terms. Certainly that has been the case through the history of the nation. Emotion and myth were important in William Hamilton Merritt's dreams of canals, Thomas Keefer's philosophizing on railroads, the nation-building imagery of CPR, C. D. Howe's aspirations for a national airline, and the Canadian fascination with the car. Everything indicates that emotion and myth will continue to be important in the future.

ENDNOTES

1. Edith Dyell, *Canada: The New Nation* (Toronto: W. Gage, no date), pp. 390–91. This was a standard textbook for Grade 7 in Ontario schools for many years in the 1950s and early 1960s.
2. A great deal has also been written about the role of geography. For a brief example see Cole Harris, "The Myth of the Land in Canadian Nationalism," in *Nationalism in Canada*, edited by Peter Russell (Toronto: McGraw-Hill Ryerson, 1966), pp. 27–46.
3. D. A. MacGibbon, *Railway Rates and the Canadian Railway Commission* (Boston: Houghton Mifflin Co., 1917), p. 54.
4. The best overview of the freight-rate controversy in Canadian history is Howard Darling, *The Politics of Freight Rates: The Railway Freight Rate Issue in Canada* (Toronto: McClelland and Stewart, 1980).
5. Edwin C. Guillet, *The Story of Canadian Roads* (Toronto: University of Toronto Press, 1966), p. 14.
6. W. J. Eccles, *The Canadian Frontier, 1534–1760* (New York: Holt Rinehart and Winston, 1969), pp. 4–5.
7. For an overview of the French regime see G. P. de T. Glazebrook, *A History of Transportation in Canada*, Vol. 1 (Toronto: McClelland and Stewart, 1964), Chap. 1.
8. *Ibid.*, Vol. 1, p. 68.
9. See Glazebrook, *A History of Transportation in Canada*, Vol. 1; Donald Creighton, *The Commercial Empire of the St. Lawrence* (Toronto: Macmillan, 1956); Doug Owram, *Building for Canadians: A History of the Department of Public Works 1840–1960* (Ottawa: Public Works Canada, 1979).
10. Glazebrook, *A History of Transportation in Canada*, Vol. 1.
11. As Konrad Studnicki-Gizbert has pointed out, Canada was, in modern terminology, a developing country in terms of transportation until the 20th century. It had no infrastructure, and yet an infrastructure was crucial to further development. Accordingly, it had to expend a large portion of available funds on construction. See Konrad Studnicki-Gizbert, "Transportation," in *The Canadians 1867–1967*, edited by R. C. Brown and J. M. S. Careless (Toronto: Macmillan, 1967), pp. 502–24.
12. H. G. Aitken, *The Welland Canal Company: A Study in Canadian Enterprise* (Cambridge: Harvard University Press, 1954), p. 86.
13. As an example see Peter Baskerville, "Entrepreneurship and the Family Compact; York-Toronto, 1822–1855," *Urban History Review*, 9 (1981), pp. 15–34. See also Owram, *Building for Canadians*, chaps. 2–3. As Creighton has shown, the situation in Lower Canada before the rebellion was more complex, with divisions along linguistic lines often impeding development. By the time of the canal-building era in the 1840s, however, a coalition of commercially minded politicians, both francophone and anglophone, had accepted the basic need for development.

14. This notion of a head-on battle with the United States for trade supremacy is best described in Creighton's classic study, *The Commercial Empire of the St. Lawrence*. See also Gilbert Tucker, *The Canadian Commercial Revolution 1845–1851* (New Haven: Yale University Press, 1936).
15. National Archives of Canada (NAC), Records of State for Upper Canada (RG 1E3), Vol. 60A, Council Chambers, January 8, 1799.
16. The protest was contained in Robert Gourlay's 1819 grievances against the Upper Canadian government. See Gerald Craig, *Upper Canada, The Formative Years 1784–1841* (Toronto: McClelland and Stewart, 1963), p. 94.
17. Speech of R. B. Sullivan, cited in *British Colonist*, September 1, 1841.
18. See A. W. Campbell, *Road Bulletin* No. 2 (Ontario Department of Agriculture, 1896) for a strong condemnation of the statutory system.
19. Glazebrook, *A History of Transportation in Canada*, Vol. 1. Chap. 4 has a short history of roads in New France and early Upper Canada. See also Guillet, *The Story of Canadian Roads*, Chap. 1.
20. Guillet, *The Story of Canadian Roads*, pp. 65–66.
21. Glazebrook, *A History of Transportation in Canada*, Vol. 1, p. 101.
22. *Ibid.*, p. 134.
23. *Ibid.*, p. 137.
24. Walter Houghton, *The Victorian Frame of Mind* (New Haven: Yale University Press, 1957); Asa Briggs, *Iron Bridge to Crystal Palace: Impact and Images of the Industrial Revolution* (London: Thames and Hudson, 1979).
25. Thomas Keefer, *The Philosophy of Railroads* with an Introduction by H. V. Nelles (Toronto: University of Toronto Press, 1972), p. 3. Original published in 1849.
26. *Ibid.*, p. 10.
27. See, as early examples, Millington Henry Synge, *Canada in 1848* (London, 1848) and F. A. Wilson and A. B. Richards, *Britain Redeemed and Canada Preserved* (London, 1850). For an American counterpart see Asa Whitney, *A Project for a Railroad to the Pacific* (New York, 1852).
28. On the railways' impact on industry see Paul Craven and Tom Traves, "Canadian Railways as Manufacturers, 1850–1880," in *Perspectives on Canadian Economic History*, edited by Douglas McCalla (Toronto: Copp Clark Pitman, 1987), pp. 118–43.
29. Kenneth Norrie and Douglas Owram, *A History of the Canadian Economy* (Toronto: Harcourt Brace Jovanovich, 1991), pp. 225–27.
30. Glazebrook, *A History of Transportation in Canada*, Vol. 1, Chap. 5; G. R. Stevens, *History of the Canadian National Railways* (New York: Macmillan, 1973); J. M. S. Careless, *The Union of the Canadas: The Growth of Canadian Institutions 1841–1857* (Toronto: McClelland and Stewart, 1967).

31. *Statutes of the Province of Canada*, 12 Vict. c. 29.
32. *Ibid.*, 14 & 15 Vict. c. 51.
33. W. L. Morton, *The Critical Years: The Union of British North America 1857–1873*.
34. G. R. Stevens, *Canadian National Railways*, Vol. 1 (Toronto: Clark Irwin, 1960), pp. 172–85.
35. Robin Winks, *Canada and the United States; The Civil War Years* (Montreal: Harvest House, 1971), Chap. 6.
36. *Statutes of Canada*, 31 Vict. (1867–1868), c. 13, *An Act respecting the construction of "The Intercolonial Railway"*; British Columbia Terms of Union, 1871; s. 11: Prince Edward Island Terms of Union, 1873.
37. Cited in Owsram, *Building for Canadians*, p. 96.
38. See Sir Sandford Fleming, *The Intercolonial* (Ottawa: Queen's Printer, 1877).
39. Ernest R. Forbes, *The Maritime Rights Movement, 1919–1927: A Study in Canadian Regionalism* (Montreal: McGill-Queen's University Press, 1979), pp. 22–23, 25–27.
40. Silver Donald Cameron, "The Iron Road to Yesterday and Tomorrow," cited in Jo Davis, *Not a Sentimental Journey: What's behind the VIA Rail cuts and What YOU can do about it* (Gunbyfield Publishing, 1990), p. 18.
41. Peter George, "Rates of Return in Railway Investment and Implications for Government Subsidization of the Canadian Pacific Railway," *Canadian Journal of Economics* 1 (1968): pp. 740–62; and his "Rates of Return of Government Subsidization of the Canadian Pacific Railway: Some Further Remarks," *Canadian Journal of Economics* 8 (1975): pp. 591–600. For a criticism of George see Lloyd J. Mercer, "Rates of Return and Government Subsidization: An Alternate View," *Canadian Journal of Economics* 6 (1973): pp. 428–37.
42. *Canadian Pacific Facts and Figures* (Canadian Pacific Foundation Library, 1946), p. 71.
43. See, as one of hundreds of examples, *Toronto Globe*, December 13, 1880.
44. *The Economist*, 39 (February 19, 1881), p. 219.
45. Advertisement in *Maclean's Magazine*, April 15, 1949, p. 5.
46. See as examples J. C. James and Alan MacDougall, *The Western Division of the Canadian Pacific Railway* (London: Institute of Civil Engineers, 1884); Harry Jones, *Railway Notes in the Northwest* (no publisher, 1884); J. H. Secretan, *Canada's Great Highway: From the First Stake to the Last Spike* (London: John Lane, The Bodley Head Ltd., 1924); Sir Sandford Fleming, *Report on Surveys and Preliminary Operations on the Canadian Pacific Railway up to January 1877* (Ottawa: McLean Roger, 1877).
47. J. P. Sheldon, *To Canada and through It with the British Association* (Ottawa, 1885).
48. R. G. MacBeth, *The Romance of the Canadian Pacific Railway* (Toronto: The Ryerson Press, 1924), p. 46.

49. Cited in John Murray Gibbon, *Steel of Empire: The Romantic History of the Canadian Pacific, the Northwest Passage of Today* (New York: Bobbs-Merrill, 1935), p. 296.
50. Historic plaque at the site of the last spike. Cited in Progressive Conservative Party, "The Last Straw: Report of the Task Force on Rail Passenger Service," 1981, p. 1.
51. Allen Smith, "The Myth of the Self-made Man in English Canada 1850-1914," *Canadian Historical Review* 59, 2 (June 1978), p. 189.
52. Gibbon, *Steel of Empire*, p. 313.
53. Walter Vaughan, *The Life and Work of Sir William Van Horne* (New York: The Century Co., 1920), p. 34.
54. Heather Gilbert, *Awakening Continent: The Life of Lord Mount Stephen, Vol. 1: 1829-91* (Aberdeen: Aberdeen University Press, 1965); D. Cruise and A. Griffiths, *Lords of the Line* (Markham: Viking, 1988); Pierre Berton, *The Last Spike: The Great Railway 1881-1885* (Toronto: McClelland and Stewart, 1971).
55. Cruise and Griffiths, *Lords of the Line*, p. xvii.
56. *Canadian Pacific Facts and Figures*, p. 10.
57. Goldwin Smith, *Canada and the Canadian Question*, with an introduction by Carl Berger (Toronto: University of Toronto Press, 1971), p. 4. Original published in 1891.
58. O. D. Skelton, "General Economic History 1867-1912," in *Canada and Its Provinces*, Vol. 9, edited by Adam Shortt and Arthur Doughty (Toronto: Glasgow Brook and Co., 1914), pp. 95-276. O. D. Skelton, *The Railway Builders: A Chronicle of Overland Highways* (Toronto, 1916).
59. Canada, *Report of the Royal Commission on Dominion-Provincial Relations* (Ottawa: King's Printer, 1940). Perhaps even more influential than the report itself were some of the studies undertaken for the Commission. See, for example, W. A. Mackintosh, *The Economic Background of Dominion-Provincial Relations* (Ottawa: King's Printer, 1939).
60. W. A. Mackintosh, "Economic Factors in Canadian History," *Canadian Historical Review* IV, 1 (1923).
61. Carl Berger, *The Writing of Canadian History: Aspects of English-Canadian Historical Writing since 1900*, second edition (Toronto: University of Toronto Press, 1986), p. 97. See also Harold A. Innis, "Transportation as a Factor in Canadian Economic History," in *Essays in Canadian Economic History*, edited by Mary Q. Innis (Toronto: University of Toronto Press, 1956), pp. 62-77.
62. See Barry G. Ferguson, "The New Political Economy and Canadian Liberal Democratic Thought: Queen's University 1890-1925" (PhD thesis, York University, 1982), p. 124 ff.
63. On nationalism and historical interpretation see Ramsay Cook, "La Survivance English-Canadian Style," in his *The Maple Leaf Forever*, 2nd edition (Toronto: The MacMillan Company of Canada Limited, 1977).
64. Berger, *The Writing of Canadian History*, pp. 241-42.

65. Creighton, *The Commercial Empire of the St. Lawrence*.
66. Donald Creighton, *John A. Macdonald: Vol. 2: The Old Chieftan* (Toronto: Macmillan, 1955).
67. Pierre Berton, *The National Dream: The Great Railway 1871–1881* (Toronto: McClelland and Stewart Limited, 1970), p. 7.
68. "Towards the Last Spike," lines 222–23 in *E. J. Pratt: Complete Poems*, edited by Sandra Dwja and R. G. Moyles (Toronto: University of Toronto Press, 1989), pp. 124–25.
69. John Murray Gibbon, *Steel of Empire*.
70. Ernest R. Forbes, *The Maritime Rights Movement*, p. 27.
71. Ontario, *Royal Commission on Transportation, 1938* (Toronto: King's Printer, 1939), p. 52.
72. Good Roads Association of the Province of Ontario, *Annual Report, Vol. II. 1894* (Toronto, 1895), p. 15.
73. See comments of Andrew Pattullo, on "Road-Making," in *Report of the Good Roads Association of the Province of Ontario* (Toronto: Warwick Bros., 1895), pp. 9–10.
74. On the Dawson route see *Annual Reports of the Department of Public Works, 1868–1873* and Owram, *Building for Canadians*, pp. 107–10. On the Yale Road see Guillet, *The Story of Canadian Roads*, pp. 91–101.
75. Morris Zaslow, *The Opening of the Canadian North 1870–1914* (Toronto: McClelland and Stewart Limited, 1971), p. 106.
76. O. D. Skelton, *General Economic History of the Dominion 1867–1912* (Toronto: Publishers' Association of Canada, 1913).
77. S. J. McLean, "National Highways Overland," in *Canada and Its Provinces*, Vol. 10, edited by Adam Shortt and Arthur Doughty, pp. 359–472.
78. G. P. de T. Glazebrook, *A History of Transportation in Canada*, Vol. II (Toronto: McClelland and Stewart Limited, 1964).
79. See *Provincial Statutes of Canada*, 14 & 15 Vict, c. 51.
80. Some of the best travelogues include J. Ewing Ritchie, *To Canada With Emigrants* (London, 1885); Douglas Sladen, *On the Cars and Off: Being a Journal of the Pilgrimage along the Queen's Highway to the East, from Halifax Nova Scotia to Victoria in Vancouver's Island* (London, 1895); B. Pullen-Burry, *From Halifax to Vancouver* (London, 1912). On British travellers in Canada generally during these years see R. G. Moyles and Doug Owram, *Imperial Dreams and Colonial Realities: British Views of Canada 1880–1914* (Toronto: University of Toronto Press, 1988).
81. See, for example, Canadian Pacific Railway (CPR), *Fishing resorts along the Canadian Pacific Railway, eastern division: where to go for trout, bass and maskinonge, and what it costs to get there* (Montreal, 1887); CPR, *Mississauga: a wonderfully interesting two hundred and seventy-five mile canoe trip through the Mississauga Forest reserve* (Montreal, 1921).

82. Figures from W. Kaye Lamb, *History of the Canadian Pacific Railway* (Toronto: Macmillan Publishing Co., Inc., 1977), Appendix, pp. 446–47.
83. For the most comprehensive study of Canadian Northern see Ted Regehr, *The Canadian Northern Railway: Pioneer Road of the Northern Prairies 1895–1918* (Toronto: Macmillan Publishing Co., Inc., 1976).
84. R. C. Brown and G. R. Cook, *Canada 1896–1921: A Nation Transformed* (Toronto: McClelland and Stewart, 1974), pp. 148–53, 200–203; John Eagle, “Sir Robert Borden and the Railway Problem in Canada Politics, 1911–1920” (PhD thesis, University of Toronto, 1972.)
85. Michael Bliss, *Northern Enterprise: Five Centuries of Canadian Business* (Toronto: McClelland and Stewart, 1987), p. 328.
86. T. D. Regehr, “The Canadian Northern Railway: The West’s Own Product,” *Canadian Historical Review* 50, 2 (June 1970).
87. Frederick A. Talbot, *The Making of a Great Canadian Railway* (Toronto: The Musson Book Company Limited, 1912), pp. 18, 26, 71.
88. Brown and Cook, *Canada 1896–1921*, pp. 200–203.
89. Canada, *Report of the Royal Commission to Inquire into Railways and Transportation in Canada* (Ottawa: King’s Printer, 1917).
90. Stevens, *History of the Canadian National Railways*, p. 277.
91. Statistics Canada, *Historical Statistics of Canada*, second edition, F.H. Leacy (ed.), 1983, Series H448–H457.
92. The actual bill creating the CNR was passed in 1919. *Statutes of Canada*, 9–10 Geo. V, c. 13.
93. Glazebrook, *A History of Transportation in Canada*, Vol. II, p. 176.
94. There is a series of timetables available in what is known as the Peel collection of the University of Alberta library. This material on routes and services draws heavily upon these as well as upon more general railway histories. Also, those books designed for the railway buff provide exceptional detail on the interior of coaches, facilities available, etc.
95. Lamb, *History of Canadian Pacific*, p. 268.
96. Canadian Pacific Timetable, 1897. Stopover for Field House.
97. Lamb, *History of Canadian Pacific*, chaps. 17, 23.
98. Number of train trips derived from *Canada Year Book*, 1919–1929.
99. Immigration in the 1930s would effectively disappear, going to as low as 5 percent of the pre-World War I figures. See Statistics Canada, *Historical Statistics of Canada*, Series A350.

100. The basic background on the 19th century tourist efforts of the Canadian Pacific is drawn from E. J. Hart, "See This World Before the Next: Tourism and the CPR," in *The CPR West*, edited by Hugh A. Dempsey (Vancouver: Douglas and McIntyre, 1984), pp. 151–69.
101. Mount Stephen House at Field, Glacier House at Glacier, Fraser Canyon House at North Bend.
102. The initial chalets at Lake Louise were very small. It was not until after the turn of the century that there was significant hotel accommodation on the shores of the lake.
103. Lamb, *History of Canadian Pacific Railway*, p. 268.
104. John Herd Thompson with Allen Seager, *Canada 1922–1939: Decades of Discord* (Toronto: McClelland and Stewart, 1985), p. 94.
105. On the railway hotels generally, see Howard Kalman, *The Railway Hotels and the Development of the Chateau Style in Canada* (Victoria: University of Victoria, 1968).
106. CPR, *Your Journey through the Canadian Rockies between Calgary and Vancouver* (1928).
107. Canadian National Railway System (CNR), *Annual Report for 1925*, p. 8.
108. CNR, *Annual Report for 1965*.
109. CPR, *Your Journey*.
110. *Canada Year Book*, 1930, p. 620.
111. *Roads and Transportation Association of Canada 1914–1974* (Ottawa: RTAC, 1974), p. 2.
112. James J. Flink, *The Car Culture* (Cambridge: MIT Press, 1975), p. 26–27.
113. *Statutes of Ontario*, 3 Edw. VII (1903), c. 27. Quebec introduced legislation in 1906. See *Statutes of Quebec*, 6 Edw. VII (1906), c. 13.
114. *Statutes of Prince Edward Island*, 3 Geo. V (1913), c. 7.
115. Flink, *The Car Culture*, p. 29 gives the average automobile price. The salary estimate comes from Statistics Canada, *Historical Statistics of Canada*, Tables E41–E48.
116. John B. Rae, *American Automobile Manufacturers: The First Forty Years* (Philadelphia: Chilton Press, 1959) discusses the growth of the auto industry in some detail.
117. Statistics Canada, *Canada Year Book*, 1933, p. 686.
118. "Address" of the Mayor of Hamilton, Canadian Good Roads Association (CGRA), *Proceedings* (1918), p. 5.
119. Flink, *The Car Culture*, p. 2.
120. Statistics Canada, *Canada Year Book*, 1933, p. 244.
121. Norrie and Oworm, *A History of the Canadian Economy*, Chap. 17.

122. Flink, *The Car Culture*, p. 35.
123. Tad Burness, *Cars of the Early Twenties* (New York: Chilton Book Company, 1968), p. 77.
124. W. H. Dandurand, "Address," CGRA, *Proceedings* (1915), p. 16.
125. Flink, *The Car Culture*, p. 39.
126. R. A. Falconer, "Address," CGRA, *Proceedings* (1915), p. 35.
127. *Canadian Motorist*, April 1929, p. 163; August 1928, p. 296; Burness, *Cars of Early Twenties*, p. 160.
128. Burness, *Cars of Early Twenties*, p. 194.
129. This commentary rests on advertisements in John D. Bondt, *Canada on Wheels* (Oberon Press, 1970); Burness, *Cars of Early Twenties*; the Ontario Motor League Publication, *Canadian Motorist* 1926–1930. For a somewhat different view see Charles Sanford, "Woman's Place in American Car Culture," in *The Automobile and American Culture*, edited by David Lewis and Laurence Goldstein (Ann Arbor: The University of Michigan Press, 1981), pp. 137–52.
130. As examples see the following articles from the *Canadian Motorist*: Bonnycastle Dale, "Sights of The Knights of The Road," June 1928, pp. 205–07; Clara Dennis, "Into the North of Cape Breton by Motor," June 1929, pp. 125–27+.
131. *Canadian Motorist* (May 1929), p. 217.
132. Mariana Valverde, *The Age of Light, Soap, and Water* (Toronto: McClelland and Stewart, 1991) gives an overview of some of the reformism of the age.
133. *Report of the Good Roads Association* (Toronto: Ontario Department of Agriculture, 1894), p. 5.
134. *Statutes of Canada 1917*, 7-8 Geo. V, c. 74. Roads and Transportation Association of Canada, *Roads and Transportation Association of Canada, 1914–1974*.
135. CGRA, *Proceedings* (1915).
136. On the 1920 meetings see CGRA, *Proceedings* (1920). For a sense of the public attention and for the details of the 1925 national meeting see the Transportation Association of Canada, *The Canadian Good Roads Association Scrapbook*, for 1925.
137. Tony Cashman, *The Alberta Motor Association: A History*, p. 1.
138. Canadian Automobile Association Headquarters, Ottawa, Binder entitled "Historical Documents, 1916–1967, Section 1916" (henceforth termed CAA Binders).
139. William Findlay, "Value of the Local Association in a Nation-Wide Highway Movement," CGRA, *Proceedings* (1920), pp. 43–46.
140. On the international background of the Motor Club concept see Margaret Anderson "Outline of the History and Background of the World-Wide Automobile Association," CAA Binders, File 1961.

141. *Statutes of the Province of Ontario*, 1901, 1 Edw. VII, c. 32; *Statutes of Quebec* 1907, 7 Edw. VII, c. 3; 1912, 2 Geo. V, c. 17, 21, 22, 23.
142. Thomas Church, "Address," *CGRA Proceedings* (1915), p. 12.
143. *Statutes of Canada*, 9-10 Geo. V (1919), c. 54. "The Canada Highways Act."
144. Ontario, *Report of the Royal Commission on Transportation*, 1938, p. 85.
145. Flink, *The Car Culture*, p. 149.
146. Alberta was followed by Manitoba in 1923; Prince Edward Island in 1924; Ontario, Quebec and British Columbia in 1925; Nova Scotia and New Brunswick in 1926; and Saskatchewan in 1929. Source: Ontario, *Report of the Royal Commission on Transportation*, p. 92.
147. Transport Association of Canada Library, "Canadian Good Roads Association, Scrapbooks, 1925."
148. Gordon Donald Campbell, "An Analysis of Highway Finance and Road User Impacts in Canada" (PhD thesis, Purdue University, 1956), pp. 97-99.
149. For a discussion of this relationship see Nancy Bryan, *More Taxes and More Traffic* (Toronto: Canadian Tax Foundation, 1972), p. 4; Saskatchewan, *Report of the Royal Commission on Taxation* (Regina: Queen's Printer, 1965), p. 34; Ontario Department of Highways, *A Plan for Ontario's Highways* (December 1956), p. 20.
150. *Statutes of Canada*, 9-10 Geo. V (1919), 13-14 Geo. V (1923), c. 4; 15-16 Geo. V (1925), c. 4.
151. Guillet, *The Story of Canadian Roads*, p. 163.
152. Previously, cars on Vancouver Island had to drive on the left side of the road; those on the mainland, on the right.
153. N. Vermilyen, President's Address, "Proceedings of the Twelfth Annual Meeting of the Ontario Good Roads Association," 10 in *Ontario Sessional Papers* XLVI (1914).
154. G. W. Tillson, "Wearing Surfaces," *CGRA, Proceedings* (1915), pp. 27-33.
155. Alberta, *Annual Report of the Department of Public Works*, 1922, p. 7; 1923, p. 8.
156. F. C. Biggs, *Report of the Accomplishments of the Highway Department of the Province of Ontario covering the Years 1920, 1921 and 1922* (Toronto: Government of Ontario, 1923), p. 3.
157. British Columbia, *Report of the Minister of Public Works, for the Fiscal Year 1928-29*, p. 6.
158. Alberta, *Annual Report of the Department of Public Works*, 1928-29, p. 15.
159. This material is drawn from the annual reports of the various provinces for the 1920s. On the specific Ontario mileage see Ontario, *Annual Report of the Department of Public Highways*, 1926.
160. Ontario, *Annual Report of the Department of Public Highways*, 1927, p. 74; British Columbia, *Report of the Minister of Public Works*, 1928-29, p. T78.

161. D. G. Wetherall and I. Kmet, *Useful Pleasures: The Shaping of Leisure in Alberta 1896–1945* (Edmonton: Alberta Department of Culture and Multiculturalism, 1990), p. 16.
162. See Ontario, *Report of the Royal Commission on Transportation*, 1938, Table A10.4 for a sense of relative rail-bus costs in the 1930s.
163. Alberta, *Annual Report of the Department of Public Works*, 1926–1930.
164. *Ottawa Journal*, April 18, 1925.
165. CNR *Annual Report for 1931*, p. 7.
166. Stevens, *History of Canadian National Railways*, p. 370.
167. Ian Drummond, *Progress without Planning: The Economic History of Ontario from Confederation to the Second World War* (Toronto: University of Toronto Press, 1987), p. 270.
168. Statistics Canada, *Canada Year Book*, 1933, p. 856.
169. Alberta, *Annual Report of the Department of Public Works*, 1933–34, p. 7; Ontario, *Annual Report of the Department of Public Highways*, 1933, p. 9.
170. Lamb, *History of the Canadian Pacific Railway*, p. 333.
171. On CNR in the 1920s see Stevens, *History of the Canadian National Railways*, chaps. 19 and 20.
172. These and other CNR numbers during the Great Depression are from CNR, *Annual Report*, 1929–1939.
173. Canada, *Report of the Royal Commission on Railways and Transportation in Canada, 1931–32* (Duff Commission) (Ottawa: King's Printer, 1932), p. 43.
174. CNR, *Annual Report for 1932*, p. 4.
175. These CPR numbers are derived from CPR, *Annual Report*, 1929–1939.
176. A. W. Currie notes that in the 1930s the annual CNR deficit was approximately equivalent to the entire revenue generated by Dominion income tax. A. W. Currie, *Canadian Transportation Economics* (Toronto: University of Toronto Press, 1967), p. 10.
177. Canada, House of Commons, *Debates*, February 7, 1933, pp. 1867–68.
178. *Statutes of Canada*, 23-24 Geo. V, c. 23, s. 3.
179. CNR, *Annual Report for 1934*, p. 3.
180. CNR, *Annual Report for 1939*, p. 4.
181. These figures derived from Statistics Canada, *Canada Year Book*, 1939–1946, and annual reports of the railways, 1939–1945.
182. Statistics Canada, *Historical Statistics of Canada*, Series T147–T162.

183. Canada, Privy Council Office, *Canadian War Orders and Regulations, 1942* (Ottawa: King's Printer, 1942); Department of Munitions and Supply, *Transit Controller*, October 31, 1942, Transit Order 3-B.
184. Canada, *Report of the Royal Commission on Transportation*, Vol. I (Ottawa: Queen's Printer, 1961), p. 46.
185. Canada, *Final Report of the Royal Commission on Canada's Economic Prospects* (Ottawa: Queen's Printer, 1957).
186. For an analysis of subsidy patterns see K. Studnicki-Gizbert, draft manuscript on Transportation Policy in Canada, Chap. 2, p. 4.
187. Currie, *Canadian Transportation Economics*, p. 15.
188. Campbell, "An Analysis of Highway Finance and Road User Imposts in Canada," p. 59.
189. Derived from Bryan, *More Taxes and More Traffic*, Table A-16. See also Michel Bérard, *Les Routes du Québec* (Quebec: Minister of Roads, 1964), p. 63.
190. There is a great deal of literature on this, especially in the United States. See, for a quick overview, William L. O'Neill, *American High: The Years of Confidence 1945-1960* (New York: The Free Press, 1986). Note especially his comments on "Automania," pp. 29-32.
191. Thomas Hine, *Populuxe* (New York: Alfred A. Knopf, 1986).
192. *Maclean's*, December 1, 1946, pp. 42-43.
193. *Newsweek*, August 4, 1952, pp. 32-33.
194. *Maclean's*, July 29, 1961, p. 3. See also, Elizabeth Macfarlan, "The Attachment to the Automobile: With Reference to Non-economic Factors Underlying this Attachment," a working paper prepared for the Royal Commission on National Passenger Transportation, November 14, 1990.
195. Toronto *Telegram*, May 3, 1954, p. 6.
196. Ontario Department of Highways, *A Plan for Ontario Highways*, p. 7.
197. Province of Quebec, *Annual Report of the Roads Department for 1946-47*, p. 11.
198. CGRA, *Proceedings* (1954), p. 8.
199. Hine, *Populuxe*, p. 56; to be fair, the combination family room-garage never really caught on. Even in an age relatively unconcerned with pollution, the combination of carbon monoxide and family living seemed to strain the love affair with the car.
200. "Report of the President," CGRA, *Proceedings of the Thirty-Fourth Convention of the Canadian Good Roads Association* (Victoria: CGRA, 1953), p. 6.
201. Bryan, *More Taxes and More Traffic*, Table 2; Statistics Canada, *Historical Statistics*, Series H197-H208 and H317-H331.

202. Perhaps the best single source for the road operations of the various provinces is in the "Roads Roundup" published each year as a part of the CGRA *Proceedings*.
203. Ontario Department of Highways, *A Plan for Ontario Highways*, p. 13.
204. *Road and Wheel*, Vol. XV, No. 1 (February 1966), p. 1.
205. See "Roads Roundup" of the CGRA, *Proceedings* (1960, 1961, 1962, 1963).
206. Province of Quebec, Minister of Highways, *Annual Report, 1946-1947*, pp. 11-13. See also Bérard, *Les Routes du Québec*, p. 49.
207. Statistics Canada, *Canada Year Book*, 1949, 1957.
208. Province of Quebec, *Annual Report*, 1959, p. 7.
209. Province of Quebec, *Annuaire statistique*, 1958, pp. 473-75.
210. J. H. Perry, "The Taxpayer's Viewpoint," CGRA, *Proceedings* (1953), p. 115.
211. Wilfred Sanders, "Assessing Public Opinion Concerning Highway Policy," CGRA *Proceedings* (1954), p. 79; Gordon Taylor, "Annual Review of the President," CGRA, *Proceedings* (1955), pp. 9-15.
212. See generally the "Roads Roundup" of the CGRA, *Proceedings* (1950-1956) and the annual reports of the departments responsible for highways in various provinces for the same years.
213. See the Symposium, "Who Shall Pay for the Roads We Need?" CGRA, *Proceedings* (1953), pp. 84-98 for comments on issue generally. For the Morrison citation see p. 96.
214. Canadian Automobile Association Library, "Brief Concerning Canadian Highways Presented to the Honourable R. H. Winters," 1955, p. 4.
215. On federal fiscal policy in these years see Robert Campbell, *The Politics of The Keynesian Experience in Canada, 1945-75* (Peterborough: Broadview Press Ltd., 1987). The post-war budget figures are on p. 228, Table VII.
216. Canada, *Dominion and Provincial Submissions and Plenary Conference Discussions*, August 6, 1945, p. 81.
217. Canada, House of Commons, *Debates*, November 21, 1949, pp. 2023-24.
218. *Ibid.*, pp. 2027-28.
219. National Archives of Canada (NAC), Records of the Department of Public Works (DPW), Accession 84-85/61, Vol. 1206, File 90-0-1, Part 3, "Address by Robert Winters, November 1951."
220. *Statutes of Canada*, 1949, 13 Geo. VI, c. 40. The road was to be constructed as a two-lane paved highway, 24 feet wide, with 10-foot shoulders and maximum grades of 6 percent. For a brief summary of the fiscal arrangements concerning the Trans-Canada see Bryan, *More Taxes and More Traffic*, pp. 135-42.

221. "Roads Roundup — Quebec," CGRA, *Proceedings* (1954), p. 64.
222. DPW, *Annual Report of Proceedings under the Trans-Canada Highway Act*, 1954, p. 1.
223. NAC, Records of DPW, Accession 84/85/61, Vol. 1206, File 90-0-1, Part 4, Robert Winters to J. T. Douglas, October 21, 1955.
224. DPW, *Annual Report Proceedings under the Trans-Canada Highway Act*, 1958–1962 gives a summary of the various amendments and details of progress.
225. *Ibid.*, 1962, p. 3.
226. For a brief history of the Roads to Resources Program see Guillet, *The Story of Canadian Roads*, Chap. 13.
227. Robert Bothwell, Ian Drummond and John English, "Canada since 1945," in *Power, Politics and Provincialism*, revised edition (Toronto: University of Toronto Press, 1989), pp. 187–88. For a good example of the lack of such concerns see Guillet's comments on the flexibility of roads. They, he said, could be built and abandoned "without loss, having paid for themselves many times over," *The Story of Canadian Roads*, p. 189.
228. Bryan, *More Taxes and More Traffic*, p. 139, Table 29.
229. Canadian Institute of Public Opinion, Gallup poll, September 5, 1962; *Gallup Report*, November 27, 1976.
230. Robert Bothwell and William Kilbourn, *C. D. Howe: A Biography* (Toronto: McClelland and Stewart, 1979), pp. 108–109. See also Studnicki-Gizbert, draft manuscript, pp. 285–94.
231. Trans-Canada Air Lines (TCA), *Annual Report for 1945*, pp. 5–7.
232. Gordon R. McGregor, *The Adolescence of an Airline* (Montreal, 1970), pp. 8–9.
233. TCA, *Annual Report for 1947*, p. 9.
234. For very brief summary of developments in TCA see C. A. Ashley, *The First Twenty-Five Years: A Study of Trans-Canada Airlines* (Toronto: Macmillan Co. of Canada, 1963).
235. Canada, Department of Transport, *Canada in the Jet Age: A Report on a Study of Department of Transport Air Services Needs 1962–1972* (Ottawa: Department of Transport Air Services, 1962), p. 25.
236. TCA, *Annual Report*, 1945–1954.
237. It is difficult to make direct comparisons, given the different types of services involved. TCA measurements indicate, roughly, that air travel cost twice as much per passenger-mile as rail travel in 1955.
238. TCA, *Annual Report 1954*, p. 9.
239. TCA, *Annual Report 1961*, p. 7.

240. There are no detailed studies of the imagery or selling techniques of the airlines. These comments are drawn from *Maclean's* and *Chatelaine* from 1945 through 1961, as well as from more limited checks of the newspapers of the period. As John Helliwell commented to me, however, the imagery of the airplane in the North is quite different. There it was much more of a lifeline to the outside world. Equally, the mythology surrounding the Northern bush pilot takes on some of the romance associated with pioneering adventures. See John Condit, *Wings Over the West: Russ Baker and the Rise of Pacific Western Airlines* (Vancouver: Harbour Publications, 1984); Tony Foster, *The Bush Pilots: A Pictorial History of a Canadian Phenomenon* (Toronto: McClelland and Stewart, 1990).
241. TCA, *Annual Report 1961*, p. 9.
242. Studnicki-Gizbert, draft manuscript, Part II, p. 100.
243. Stevens, *History of the Canadian National Railways*, pp. 437–38.
244. For a detailed study of the freight-rate issue — one quite critical of the way it has evolved — see Howard Darling, *The Politics of Freight Rates: The Railway Freight Rate Issue in Canada* (Toronto: McClelland and Stewart, 1980).
245. D. J. Hall, "The Manitoba Grain Act: An Agrarian Magna Charta?" *Prairie Forum* 4, No. 1, pp. 105–20; Gerald Friesen, *The Canadian Prairies: A History* (Toronto: University of Toronto Press, 1984), chaps. 9 and 10.
246. See Canada, *Report of the Royal Commission on Maritime Claims* (Duncan Commission) (Ottawa: King's Printer, 1927).
247. Forbes, *The Maritime Rights Movement, 1919–1927*, pp. 25–27.
248. Province of Nova Scotia, *A Submission of Claim with Regard to Maritime Disabilities within Confederation* (1925), p. 53.
249. *Statutes of Canada*, 17 Geo. V, c. 44.
250. *CNR vs. CPR et al.* 39 CRC 1, 25–7. Cited in Canada, *Report of the Royal Commission on Transportation*, 1951, p. 24.
251. Currie, *Canadian Transportation Economics*, p. 298.
252. Darling, *The Politics of Freight Rates*, p. 185.
253. Canada, *Report of the Royal Commission on Transportation*, 1951, p. 35.
254. See as an example Province of Manitoba, *Manitoba's Submissions to The Royal Commission on Transportation* (Winnipeg, 1949), pp. 19–28.
255. Darling, *The Politics of Freight Rates*, Chap. 9.
256. Canada, *Report of the Royal Commission on Transportation* 1951, p. 137.
257. Currie, *Canadian Transportation Economics*, p. 15.
258. Canada, *Report of the Royal Commission on Transportation*, Vol. I (1961), p. 58.

259. Canada, Royal Commission on Transportation, *Submissions* (1961).
260. *Ibid.*, Vol. 4, "Submission of Canadian National Railway," p. 5.
261. Canada, *Report of the Royal Commission on Transportation* 1961, Vol. I, pp. 11, 44. See also F. W. Anderson, "The Philosophy of the MacPherson Royal Commission and the National Transportation Act: A Retrospective Essay," in K. W. Studnicki-Gizbert, *Issues in Canadian Transportation Policy* (Toronto: Macmillan of Canada, 1974), pp. 47–75.
262. Canada, *Report of the Royal Commission on Transportation* (1961), Vol. I, p. 60.
263. On the background of the *National Transportation Act*, see Canada, House of Commons, *Debates*, 1966, pp. 7988–98. Also, Interview with J. W. Pickersgill, March 7, 1991.
264. *Statutes of Canada*, 14-15-16 Eliz. II, c. 69.
265. *Ibid.*, s. 314J deals with passenger service reductions. On the whole question of the relationship between the NTA and the MacPherson Commission, see Anderson, *Philosophy of the MacPherson Royal Commission*, pp. 47–75.
266. *Statutes of Canada*, 14-15-16 Eliz. II, c. 69, s. 314J.
267. J. W. Pickersgill, who headed the CTC through part of this period, said that the Commission took a cautious approach to abandonment partly to prevent continual recourse to appeals to the Governor General in Council. Interview with J. W. Pickersgill, March 7, 1991.
268. Canada, House of Commons, *Debates*, September 2, 1966, p. 8038 and December 21, 1966, p. 11490 provide examples of accusations that railways were deliberately running passenger service into the ground. The Minister of Transport from 1964 to 1967 felt that the railways were sincerely making an effort in the passenger area. Interview with J. W. Pickersgill, March 7, 1991.
269. CNR, *Annual Report for 1965*, p. 10.
270. Note that as of 1970 commuters on the GO system were added to railway passenger statistics.
271. Norrie and Owrap, *A History of the Canadian Economy*, Chap. 22.
272. Canadian Transport Commission (CTC), Research Branch, *Productivity Analysis of the Canadian Airline Industry*, working paper prepared by P. S. Dhruvarajan and R. F. Harris, pp. 67–69. Air Canada, *Annual Report 1972–1976*.
273. Ralph Nader, *Unsafe at Any Speed: The Designed-in Dangers of the American Automobile* (New York: Grossman Publishers, 1965).
274. Darryl Newbury, *Stop Spadina: Citizens Against an Expressway* (Mississauga: Commonact Press, 1989); *Vancouver's Transportation Future* (Simon Fraser University, Geography Dept., 1972).
275. Robert Bott, David Brooks and John Robinson, *Life After Oil: A Renewable Energy Policy for Canada* (Edmonton: Hurtig, 1983).

276. G. Leach, "Energy and the Car," in Meredith Thring, *Mankind and the Engineer*, Vol. 2 (Herts: Peregrinuin, 1974), p. 108.
277. Energy, Mines and Resources Canada, *Energy Futures for Canadians* (Ottawa: Supply and Services, 1978), p. 97.
278. Transport Canada, *An Interim Report on Inter-City Passenger Movement in Canada* (Ottawa: Department of Transport, 1975), p. 77.
279. See, for example, Science Council of Canada, *Canada as a Conserver Society*, Report 27, (Ottawa: Supply and Services, 1977) or *Transportation in a Resource Conscious Future: Intercity Passenger Travel in Canada*, Report 34 (Ottawa: Supply and Services Canada, 1982).
280. On the changing perspective in the United States see Robert Ackerstone, "Some Milestones of Automotive Literature," in *The Automobile and American Culture*, edited by David Lewis and Laurence Goldstein (Ann Arbor: University of Michigan, 1980), pp. 394-404.
281. Ontario, *Annual Report of the Department of Highways*, 1971, p. 1.
282. "President's Address," CGRA, *Proceedings* (1970), p. 4.
283. The DBS Series was 53-201, "Road Mileage and Street Expenditures."
284. On GO see Ontario, *Annual Report of the Department of Highways*, 1969, p. 1.
285. *Road and Wheel*, XV, 1 (February 1966), p. 1; XVI, 1 (February 1967), p. 1.
286. On 1980s see Council of Ministers Responsible for Transportation and Highway Safety, *National Highway Policy for Canada, Phase 2*, Table TA.
287. Canada, *Report of the Royal Commission on Transportation*, Vol. 1 (1961), pp. 26-28.
288. Michael Jackson, "Introduction," in *Proceedings of the First National Rail Passenger Conference*, edited by Michael Jackson (University of Regina, Department of Extension, 1977), pp. 4-5.
289. Ibid., "Address by Hon. Otto Lang," p. 16.
290. Interview with Otto Lang, Winnipeg, March 5, 1991.
291. Howard J. Darling, "An Historical Review of Direct Transport Subsidies in Canada," a report commissioned by the Canadian Transport Commission (June 1975). See also CTC Research Branch, *A Review of the Existing Intercity Passenger Transport Systems* (1975), Table 8.2, which gives a route by route breakdown of the subsidies as they were in the early 1970s.
292. Transport Canada, *An Interim Report on Inter-City Passenger Movement in Canada*, p. 115.
293. Canada, House of Commons, *Debates*, September 1, 1966, p. 7989.
294. See CTC, *Annual Report 1970-1973* for summaries of decisions in this area. In discussing the CTC policy in these years Jack Pickersgill, as head of the CTC, believed that that body should not get ahead of public or political opinion and thereby face possible reversal of

- its decisions and undermining of its credibility. Interview with J. W. Pickersgill, March 7, 1991. See also CTC, Research Branch, *An Analysis of Railway Transport Committee Decisions 1967–1980* (Ottawa: CTC, 1982).
295. Canada, House of Commons, *Debates*, January 29, 1976, pp. 10437–47.
296. For the directives to the CTC see Transport Canada, *Highlights of Rail Passenger Policy Announcement by Transport Minister Otto Lang*, January 29, 1976.
297. Canada, House of Commons, *Debates*, January 29, 1976, p. 10438. Interview with Otto Lang, March 5, 1991.
298. CTC, Railway Transport Committee, *Final Plan for Eastern Transcontinental Passenger Train Service* (June 1977); *Preferred Plan for Western Transcontinental Passenger Train Service* (April 1977). See, for press coverage, *Globe and Mail*, February 15, 1977, p. B1; April 18, 1977, p. 6; May 3, 1977, p. B1; May 28, 1977; *Winnipeg Free Press*, March 22, 1977, p. 6; May 18, 1977, p. 7.
299. Vote 52d Department of Transport *Appropriation Act No. 1*, 1977.
300. Interview with Harry Gow, President of Transport 2000, January 15, 1991.
301. For a sense of the attention, see the *Halifax Chronicle-Herald's* reporting of the CTC, Rail Committee hearings in May 1977. *Halifax Chronicle-Herald* May 16, 1977, p. 1; "Nova Scotia notebook," May 17, 1977, p. 6; "Funds key to rail's survival," and "Service should be tied to usage — MP," May 19, 1977, p. 1; "Real study needed on transportation," May 20, 1977, p. 6.
302. VIA Rail Canada Inc., *Annual Report 1985*, p. 10.
303. For a few examples see, Science Council of Canada, *Transportation in a Resource-Conscious Future*; CTC, Research Branch; David McQueen, "Aspects of Rail Passenger Policy in Canada," memo, unpublished, 1984.
304. Science Council of Canada, *Transportation in a Resource-Conscious Future*, p. 57.
305. On the opposition to cutbacks see the testimony in Progressive Conservative Party, *The Last Straw*. On the Gallup figures see *Gallup Report*, October 24, 1981, which showed that 51 percent of Canadians disapproved of cutbacks in rail passenger service and only 27 percent approved.
306. Progressive Conservative Party, *The Last Straw*.
307. Ontario Task Force on Provincial Rail Policy, *The Future of Rail* (Toronto, 1981) pp. 43–44.
308. Statistics Canada, *Rail in Canada 1987* (Ottawa: Statistics Canada, 1989), p. 24.
309. Transport Canada, *Freedom to Move: A Framework for Transportation Reform* (July 1985), p. 1.
310. "Will there be a future for Via Rail?" *Globe and Mail*, January 19, 1989, p. A6.

311. See, as examples of public concern, *Montreal Gazette*, April 13, 1989, p. A1, "81% tell Gallup they want Ottawa to retain VIA"; April 14, 1989, p. D3, "VIA needs more help, not less"; *Globe and Mail*, p. A5, "Howls of protest over cuts in Atlantic Canada"; *Winnipeg Free Press*, October 5, 1989, p. 1, "VIA cuts greeted with tears, cheers" and p. 6 "Much pain; little gain"; *Halifax Chronicle Herald*, October 5, 1987, p. B1, "VIA cuts met with fear, anger" and "Passengers ponder life without VIA."
312. Statistics Canada, Catalogue No. 51-206 and 51-002.
313. JoAnne Bonnyman and Robert Dodd, "A Review of Low Priced Air Fares 1979" (Ottawa: CTC, Passenger and Aviation Economics, Research Branch, November 1980).
314. It is interesting that neither J. W. Pickersgill nor Otto Lang, when interviewed as part of this project, were very optimistic about the survival of the two airlines. Interviews of March 5, 1991 and March 7, 1991.
315. Roads and Transportation Association of Canada, *Roadway Infrastructure Study* (Ottawa: RTAC, 1987).
316. *Gallup Report*, October 24, 1981. It should be noted that opinion in Quebec was very close (41 percent to 40 percent).
317. *Gallup Report*, April 13, 1989; *Winnipeg Free Press*, July 10, 1989, p. 28.
318. "Polls show travel sacred to Canadians", *Winnipeg Free Press*, July 10, 1989, p. 28.
319. See, most recently, the comments of the public before the Royal Commission on National Passenger Transportation itself. Canada, Royal Commission on National Passenger Transportation, *Getting There: The Interim Report of the Royal Commission on National Passenger Transportation* (Ottawa: Supply and Services Canada, 1991), pp. 133-34.
320. VIA Rail Canada Inc., *Annual Report 1985*, p. 9.
321. Transport 2000, "Submission to the Royal Commission on National Passenger Transportation," (1991) p. 4.
322. "VIA Rail: national dream fades," *Toronto Star*, March 28, 1989, p. A4.
323. Jean-Paul Desbiens, "Trains: to people and to the land, a heritage. He remembers them well," *Globe and Mail*, May 7, 1977, p. 6.
324. Canada, *Report of the Royal Commission on Transportation* (1961), p. 32.
325. Canada, Royal Commission on Transportation, *Submissions*, Vol. 30, p. 4416; Vol. 33, p. 5184.
326. For a good example of the arguments along these lines see testimony in Progressive Conservative Party, *The Last Straw*.
327. Transport 2000, "Brief to the Railway Transport Committee at Maritime Rail Passenger Hearings, April, 1977," p. 6.
328. Transport 2000, "Inter-City Railway Passenger Transportation in Central Canada over the Next Ten Years: A Public Interest Approach to Development" (September, 1979), p. 2.

329. Canada, House of Commons, Standing Committee on Transport, *Report to the House*, November 8, 1989, p. 3.
330. "Count the social costs too," *Globe and Mail*, April 18, 1977, p. A6.
331. This unscientific observation by the author came during a March visit to Jasper. The couple had travelled from Montreal to the mountains.
332. Canada, Royal Commission on Transportation, "Hearings At Fredericton," *Submissions*, Vol. 3, November 9, 1959. Both the official delegation from the city and the Chamber of Commerce looked at air travel and concentrated on adequate direct routes rather than the particular mode of travel.

U.S. INTERCITY PASSENGER TRANSPORTATION POLICY, 1930-1991: AN INTERPRETIVE ESSAY

George W. Wilson*

October 1991

1. INTRODUCTION

The U.S. has not had anything resembling a rational, coherent or unified policy with respect to transportation in general and even less of one regarding passenger transportation in particular. Legislative changes since the late 1970s in the direction of "greater reliance upon market forces" represent the beginning of a more coherent and consistent approach but there is no certainty that this will persist. The emphasis upon some variant of the economists' efficiency criteria may turn out to be short-lived. Certainly the "economist's hour," as it has been called,¹ took a long time coming to the area of transportation policy and despite some recent successes constantly needs to be addressed, as will be noted later. Thus the characterization of the past U.S. federal government's approach to transportation policy as "A chaotic patchwork of inconsistent and often obsolete legislation [that] has evolved from a history of specific actions addressed to specific problems of specific industries at specific times" is not an unfair appraisal, all the more telling since it came from the Special Message to the Congress on Transportation by President Kennedy in 1962.² As late as 1979 and despite some 20 or more years of strong pressures from the economic community, backed up more or less by every president from Kennedy on, a major policy

* Indiana University.

study³ could with justification still assert that the government's approach to transportation remained "an assortment of policies and programs which have been developed in an *ad hoc* fashion to achieve sundry goals or resolve various issues."⁴

This is not, of course, unique to the U.S. All countries suffer from what may be referred to as the "Grand Transportation Mystique" (GTM), which identifies transportation investments and policies as having special attributes for stimulating economic growth or solving problems in other industries. They accomplish such goals by providing lower, preferential fares or rates, expanded capacity, improved service quality and so on without regard to costs or alternative investments or policies that might often prove to be more effective and less costly. There is a belief that transport should be used as an instrument for accomplishing objectives in other segments of the economy or in the "national interest."⁵ The GTM is thus a mind-set that leads governments into actions towards the transport sector that more careful thought and better knowledge would indicate to be inefficient and perhaps even counter-productive. These actions take many forms and there are numerous examples, especially in North America. But everywhere these involve enormous resources to preserve parts of a system no longer needed, to expand investment into areas of limited economic potential far ahead of potential need or demand, or to maintain a pattern of regulation that involves considerable economic waste for purposes that become increasingly difficult to explain and more difficult to justify. The GTM also encourages efforts to regulate, subsidize, operate or otherwise dabble in virtually all aspects of the provision of transport services.

To be sure, from time to time the apparent dabbling and/or expenditure may have seemed sound, perhaps sensible, necessary or maybe even "efficient." But as the economy or circumstances change, the consequences of such activities become increasingly dubious, and very costly to maintain relative to any realized benefits. The belated recognition of such past "mistakes" then gives rise to a more or less major "policy change." This study reflects upon the changes that have occurred along these lines in the U.S. with special reference to intercity passenger transportation since the late 1920s.

In Canada the major transportation policy change was identified in the MacPherson Report⁶ which led to new legislation in 1967. The new policy emphasized the workability of competitive forces in Canada, especially in

the post-war environment, and the resultant desirability of refraining from using any particular mode of transport, especially railroads, as an instrument of national policy. Greater reliance upon market forces was to be counted on in the interests of efficiency and, in general, economic criteria were to be emphasized and regulation minimized. Where subsidy of any kind seemed desirable to accomplish some specific public purpose, the railroads were to be compensated for the losses involved. The previous "monopoly" position treated such loss-generating activities as a kind of *quid pro quo* for past government support and which could then be readily funded by cross-subsidy from so-called captive traffic. Such internal subsidization was no longer feasible under competition and never very economically efficient.

While the MacPherson Report focussed mainly on railroads and freight, it encompassed all modes. As the Report itself noted, the "freight rate inequity problem" was its "principal raison d'être."⁷ Yet the emphasis upon the pervasiveness of competition in transportation certainly was believed applicable to passenger transportation as well. Rail passenger deficits were uneconomic and to be eliminated as the public had indicated its preference for other modes of travel.⁸ Intercity passenger travel was mainly an after-thought or ignored altogether in most major U.S. transportation reports that I have been asked to examine (see Table 1) except for the adverse impacts of passenger deficits on the rail freight business. But the MacPherson Report stands out among its would-be U.S. counterparts as having taken the lead in establishing a major policy shift towards the totality of commercially provided transportation. Not a single U.S. report was so far ahead of the general economic philosophy of the government at the time. In fact, MacPherson led to a consensus in favour of change in Canada and paved the way for major legislative change within six years.

Similar sentiments were brewing at the same time in the U.S. but no study or report succeeded in capturing the flavour of what needed to be done with sufficient persuasiveness or power to culminate in implementing legislation. Indeed, the requisite legislation did not even begin in the U.S. until the late 1970s and early 1980s. The various U.S. reports, studies and the like mainly reflected a growing consensus among those interested in transportation (including academics, politicians, lawyers, regulators, providers of transportation services, shippers as well as intelligent and concerned lay-people) that existing arrangements were becoming increasingly counter-productive and that alternatives were needed. References to U.S. reports

in this study are mainly included to highlight the slow and tepid approach to substantive revision of U.S. transport policy compared with Canada. In only one major aspect was the U.S. ahead of Canada; Amtrak was created in 1970 while VIA Rail was delayed seven years, hardly something to be proud of under a transport philosophy stressing market forces and elimination of most non self-support services.

Table 1

MAJOR U.S. TRANSPORTATION POLICY REPORTS AND STATEMENTS

1. U.S. Department of Commerce, Office of the Federal Coordinator of Transportation, *Passenger Traffic Report* (Washington, D.C.: Government Printing Office, 1935). And various other reports of the Federal Coordinator.
2. U.S. Department of Commerce, *Modern Transport Policy; Documents Relating to the Report of the Presidential Advisory Committee on Transport Policy and Organization and Implementing Legislation* (Washington, D.C.: Government Printing Office, 1956). (The Weeks Report.)
3. Interstate Commerce Commission, *Railroad Passenger Train Deficit*, Report prepared by Howard Hosmer, Hearing Examiner, et al., ICC Docket No. 31954, 1959. (The Hosmer Report.)
4. U.S. Senate, Special Study Group on Transportation Policies in the United States. *The Doyle Report: National Transportation Policy*, (Washington, D.C.: Government Printing Office, 1961).
5. Special Message to the Congress on Transportation, by President Kennedy, The White House, April 5, 1962.
6. U.S. Department of Transportation, U.S. Federal Railroad Administration, *Report on the Potential for Integrating Rail Service Provided by the National Railroad Passenger Corporation with Other Modes* (Washington, D.C.: Government Printing Office, 1976).
7. National Transportation Policy Study Commission, *National Transportation Policies Through the Year 2000* (Washington, D.C.: Government Printing Office, 1979).
8. U.S. Department of Transportation, *Moving America* (Washington, D.C.: Government Printing Office, 1990).

THE SUBMERGENCE OF EFFICIENCY CRITERIA

It is not immediately obvious why the U.S. delayed so long in implementing the kind of change in policy that Canada had accomplished by 1967. The Canadian environment was certainly less structurally competitive than that of the U.S. With only two major railroads and air carriers, a less ubiquitous highway system than the U.S. and generally smaller domestic markets, competition in Canadian transportation would have appeared to be less

workable. Given Canada's lesser emphasis upon market orientation, or "capitalism" in general, the move to deregulate a substantial aspect of the economy was a bold one that might have been expected first from the U.S., especially in an area deemed as important and historically significant as transportation. Certainly U.S. economists did not lag behind economists in Canada in their analysis of the need and enthusiasm for transport regulatory reform. Many even testified before the MacPherson Commission hearings and strongly supported the final report. Most loudly cheered the passage of the *National Transportation Act of 1967* and widely proclaimed that Canada had indeed taken the lead in developing sensible economic policy toward transportation, a lead that the U.S. should promptly follow. However, more than a decade was to elapse before the U.S. got around to serious policy revisions. Even then they were partial and modal.

Thus one of the themes to be explored in this essay is why the long delay and more important, why ad hocery with respect to transportation policy persisted for so long in the U.S. It was surely not too radical to adopt an alternative that emphasized efficiency and market orientation or at least an approach that sought to assess the benefits vis-à-vis the costs of alternative and existing policies much earlier, especially in the bastion of *laissez-faire*. Certainly the virtual ignoring of efficiency criteria prior to the 1960s in Canada and the late 1970s in the U.S., as well as the earlier timing of the pro-competitive policy in Canada, cries out for explanation. The almost exclusive concern in U.S. transportation policy has been for something vaguely defined as the "national interest" which had a heavy dose of defence in it, to say nothing of "regional development or balance." Both of these notions are deliciously vague but when mixed with emphases upon "equity," "justice," "reasonableness," "fairness," "adequacy" and so on, they defy conceptual or empirical clarity. Since the goals of transport policy were never clearly spelled out, the ingredients of policy could not be related to any purpose.

Thus for example, the first explicit statement of national transportation policy in the U.S., which appeared in the *Transportation Act of 1940* is a welter of ambiguity. It reads as follows:

It is hereby declared to be the national transportation policy . . . to provide for fair and impartial regulation of all modes of transportation subject to the provisions of this Act, so administered as to recognize

and preserve the inherent advantages of each; to promote safe, adequate, economical and efficient service and foster sound economic conditions in transportation and among the several carriers; to encourage the establishment and maintenance of reasonable charges for transportation services, without unjust discriminations, undue preferences or advantages, or unfair or destructive competitive practices; to cooperate with the several States and the duly authorized officials thereof; and to encourage fair wages and equitable working conditions; — all to the end of developing, coordinating, and preserving a national transportation system by water, highway, and rail, as well as other means, adequate to meet the needs of the commerce of the United States, of the Postal Service, and of the national defense. All of the provisions of this Act shall be administered and enforced with a view to carrying out the above declaration of policy. ('54 Stat. 899, 1940.)

To be sure, all policy statements are vague and general. This one at least mentions the word "efficient" but it is buried or lost amid notions of "just," "reasonable," "undue," "adequate," etc.

Under this rubric and even earlier, the Interstate Commerce Commission (ICC) for example, concerned itself about the stability of the railroad rate structure, about the need to preserve "common carriage," about the need to prevent any regulated mode of transport from driving another mode out of business or even one common carrier "injuring" another. Common carriers were to "share" markets, not compete against each other, in the search for "preservation of inherent advantages." Sometimes these *desiderata* were invoked for purposes of national defence,⁹ sometimes for the needs of shippers and passengers or for something else entirely such as "stability." Transport regulation often vacillated between protecting the public and protecting the firms regulated. Too often many of the markets were preserved for a selected number of regulated common carriers thereby creating monopoly or oligopoly — like situations which were then used as justification for continued regulation. The entire airline passenger industry was effectively cartelized by the Civil Aeronautics Board (CAB) for 40 years after its formation without consideration of its competitive possibilities or in the naive belief that all transport competition was "destructive." The courts occasionally reminded the regulatory commissions that they were not given carte blanche authority to create a cartel out of a competitive situation.

At the same time, no regulatory agency was able to ensure that carrier costs were minimized for any level of output because there was no way to determine the costs or quantities of particular inputs, the technology employed or managerial aptitudes. The cost side of carriers' operations were beyond the control of regulation. In the absence of competitive pressures and with the inability to compete on a price basis, regulated transportation was inefficiently provided. If we add to this the "other" costs of regulating, such as the deliberate creation of excess capacity (for defence purposes, of course), the costs of running the regulatory agency itself, the wastes involved in the innumerable and lengthy cases designed to give "due process" to all whether or not directly involved, the diversion of managerial talent from running the business effectively to taking care of matters of regulation and so on, it is clear that the regulatory aspect of public policy, so pervasive but uneven in transportation, pretty much ignored efficiency of production.

Since the other main aspect of public policy, namely provision or financing of infrastructure, was also bereft of sensible investment or benefit-cost criteria, it is evident that efficiency considerations got short shrift. In fact, anti-efficiency sentiment was so strong that attempts to make public investment in transportation subject to such criteria were systematically expunged from any proposed legislation. Nor has there been much success in developing an efficient set of user-charges for publicly provided rights of way.

Thus while the major thrust in the U.S. over the past decade has been in the direction of market forces and efficiency, there is considerable distance yet to go in terms of public investments, user-charges and even adequate enforcement of antitrust laws to forestall anti-competitive mergers and practices. But the approach to change in transportation policy and the more overt economic goals suggest at least a conceptual clarity. There is now a greater willingness to view transportation as analogous to other industries with no unique aura or magic entitling it to special attention. The GTM has been tarnished at least. It is, and ought to be, subject to the regular impacts of supply, demand, competition, technology and so on affecting all productive aspects of the economy and judged by similar criteria. The prime goal of transportation policy should, therefore, be "Efficiency within transportation, both in terms of use of existing infrastructure, vehicles, and rolling stock . . . and in terms of investment criteria . . . deviations from which require special justification. . . ." ¹⁰

THE DISTINCTION BETWEEN PASSENGER AND FREIGHT TRANSPORTATION

Throughout the changes in U.S. transportation policies the emphasis has almost completely been on freight transportation. The original *Act to Regulate Commerce*, 1887 was the result of shippers' complaints of excessive rate discrimination; although written with sole regard to freight, it was made applicable to rail passenger traffic as well simply because rails provided both services. Bus service was included in the *Motor Carrier Act of 1935* virtually as an afterthought. Relatively few cases or proceedings were held in connection with bus transportation compared with intercity truck commodity movements. Even the creation of Amtrak was less a direct passenger-oriented measure than a means of bailing out the railway freight industry from ever-increasing passenger deficits. Almost all changes affecting passenger traffic have come about through attempts to respond to problems originating in the movement of freight. This is true of the MacPherson Report as well.

Joseph B. Eastman, one of the most important American regulatory commissioners ever, plaintively remarked in 1936 that "passenger service should not be considered and dealt with as an unavoidable but unwanted stepchild of railway operation."¹¹ Yet as far as *commercial* passenger service is concerned, aside from airline policy, passenger traffic has received little overt attention. This should not necessarily be lamented. Many of the carriers receiving most of the regulatory attention often wished they had been less "favoured." Even the airlines, overwhelmingly a passenger mode, received their early subsidies in the U.S. not on the basis of the number of passengers or passenger miles, but as "mail-pay" (that is, goods movement) contracts. The subsequent deregulation was based upon earlier arguments made for freight markets in general and preceded by prior deregulation of air cargo. The statement of national transportation policy in 1940 noted above omits explicit reference to air transportation probably because the latter deals mainly with passenger traffic. As one of the few texts on passenger transport puts it, "the subject of this declaration is really freight transportation."¹² In fact, air regulation was at one time to be placed under ICC jurisdiction but fear of overloading the Commission already swamped by being given jurisdiction over motor carriers in 1935,¹³ led to creation of a separate commission which later became the Civil Aeronautics Board.

The assumption underlying the lack of distinction between passenger and freight transportation is that the differences between moving people rather than goods are not important. Furthermore the relative indifference regarding commercial passenger travel from a policy perspective is doubtless related to its small amount compared with total intercity passenger traffic. Passenger revenues were never a very large proportion of total rail revenues and ever since the ICC kept the data, from 1936 until the creation of Amtrak, rail passenger traffic, leech-like, recorded deficits which were made up from net revenues from goods movement. The massive highway programs of the U.S. and the concomitant development of the private automobile have relegated the size of the entire commercial intercity passenger market to somewhere between 10 and 15 percent of the total. The vast majority of the populace have feasible travel options somewhat independent of air but obviously independent of bus and rail, in the sense that, if the latter two simply disappeared, the *national* inconvenience would be slight.

What are the differences between passenger and freight transportation and do they matter as far as public policy is concerned?

There are some obvious differences between people and goods transportation. Service quality is more important for passengers than freight. Comfort of the ride is qualitatively different and more significant for people. It may be an overstatement but there is some truth in the assertion that "getting there is half the fun." Passengers have to be individually placed and guided to often pre-arranged seats unlike tons of coal or even packages of freight, and then must be more or less "entertained" *en route*. The terminal facilities need to be of considerably higher quality for people. Whether they had to be the opulent structures of many of the old railway passenger terminals which became landmarks in many parts of the country, a kind of North American substitute for European castles, is not clear. They must however be especially well organized to move people on and off vehicles or aircraft with a degree of convenience, care and dispatch unnecessary for goods. Baggage handling is another function that occasions special handling, extra costs and more careful organization. Unlike freight, passengers can also complain. Although freight shippers often voice their opinions of the service, there are fewer of them than there are passengers, and the complaints are likely to be of a different sort. The quality dimension of passenger service is thus likely to loom as far more important and of a different kind than for commodities.

All this suggests that passenger cost functions are likely to differ from those of freight depending upon the mode — air, rail, bus or water. To be sure, analytically there is no reason why a cost function related to an output unit called ton-miles rather than passenger-miles should differ in general shape. Indeed, economic analysis, in seeking to be general, makes no distinction among the various kinds of output or output units and remains wedded to a theory that leaves the output unit carefully undefined and unspecified. It thus makes no difference whether output is something physical (such as tons of steel, bushels of wheat, units of homogeneous widgets, etc.) or some non-storable service, (such as transportation, entertainment, medical treatment, etc.).

If no distinction is made between physical and service units in general economic analysis, it is scarcely to be expected that passenger-miles and ton-miles would get separate treatment although neither of these often-used transport measures can be construed as homogeneous. That is another and longer story that I have dealt with elsewhere and which needs no elaboration here.¹⁴ Suffice to say that a proper conception of output units is essential not only for empirical research on both cost and demand functions but to pin-point that elusive but important concept in economics, *marginal cost*. We need to know marginal cost of *what* for pricing and costing purposes and for different qualities of any quantitative output unit. Fortunately much progress has been made along these lines.¹⁵ But the lack of widespread concern over output units has made it easier to ignore the distinction between moving things and people. Both involve "transport" and for many that is good enough.¹⁶

Furthermore the nation's estimated freight bill is vastly in excess of the estimated passenger bill if we exclude private automobiles. Some estimates suggest that the freight bill is some five and a half times larger than the passenger bill and that this has remained fairly stable over the years.¹⁷ Thus the insignificance of commercial passenger travel relative to freight movements in the U.S. may account for much of the neglect of the former and would partly explain the concentration of regulatory policy changes over time on the freight aspects of transportation. People-movement is overwhelmingly by private car and not for profit and thus not subject to the kinds of economic or business constraints oriented to competition control, profit regulation and route, back haul and other restrictions. Indeed, freedom of travel along publicly provided rights of way in one's own

vehicle is taken for granted. Automobile ownership is now virtually universal in the U.S. Yet this was not always the case and during earlier periods, say the 19th century, freight still ruled the aim, end and process of policy formation.

Hansen has remarked that the "distinction between freight and passenger transportation was often obscured" but that in the "developmental context" this was of little importance. In the 19th century, "Freight and passenger traffic would thus both have to be served . . . and railroads, turnpikes and waterways carried significant amounts of each. It was common to lump both together by viewing these transportation facilities as 'means of communication'." ¹⁸ Thus the opening up of the west, for example, required both people and commodities to move back and forth. Settlers could be herded into freight trains and "shipped" west like other "commodities," including settlers' effects. In fact, many were.

The foregoing suggest several reasons why the distinction is more significant than it might have seemed in the past. One is that the safety issue now looms larger. The difference between loss and damage claims for commodities is of far less concern than death and injury to passengers. In a nation spending vast amounts to improve the health and safety of its citizens, death and injury from transportation of people has taken on new importance. This of course includes private automobile travel as well. But concern for such issues has risen from a fatalistic "acts-of-God" notion to one that society has created by its own technological advances in transportation and elsewhere. It is now incumbent to direct resources to alleviate such hazards to the extent possible. Highway and airway safety concerns have taken on new dimensions as new technologies both create more potentially dangerous situations and provide increasing opportunities for remedial or preventive actions. These include the so-called "smart" cars and highways, improved aircraft guidance and control systems and soon, stricter safety enforcement measures and the concomitant rise of damage claims for safety violations and accidents. In short, the provision of passenger service by any firm is attendant with concerns for the safety of the passengers that is greater than those in the case of freight.

In addition, comfort has now become far more significant if only because the private car provides a convenient reference point for both comfort and convenience that is hard for a commercial carrier to match. A successful

commercial operation requires a degree of concern and ability to address these matters which are highly personal, that no freight agent ever had to contend with.

This suggests several things. Since the cost and demand elements are so different in the markets for freight and passenger intercity transportation and since societal rules of the game, so to speak, have shifted substantially in the direction of human safety and comfort, and other quality dimensions, the capabilities of those providing passenger services are significantly different from those who transport goods. What is an adequate railroad bed for freight traffic will no longer suffice for the smooth ride demanded by passengers accustomed to smooth highways or airways. Indeed new transport technologies, such as maglev need their own distinct right of way. It seems clear that rail passenger traffic will require separate, differentially maintained and perhaps unique rights of way if railways are to carry people. More generally, the present trend toward separate management of goods and people movement will probably continue. Air, truck, rail and water cargo carriers will increasingly involve companies totally divorced from significant passenger movements and vice versa. Amtrak and the other proposals for high-speed rail passenger intercity travel involve self-owned rights of way and separate corporations or ownership. No longer can one expect a single railroad to move relatively large numbers of people in a predominantly freight business.

The 19th century, North American model of development, which had to cope with substantial empty spaces required both people and goods, as Hansen has noted. People were needed for labour to produce a surplus and to create demands for production materials and consumption; goods were needed for export and import for both purposes. Thus a two-way, in some sense, equivalent movement of people and goods was crucial. The catalyst in all this involved some means of conveyance and rights of way. The better these were, the faster the development, or so the story went in Canada and the U.S. Probably this was the origin of the Grand Transportation Mystique. However, now that there are far fewer empty economic or geographic spaces, the people movement aspect of the developmental model drops relative to inter-regional goods movement associated with the geographic specialization already in place. There is still much people mobility but less associated with permanent settlement or even business than was the case before, say World War II. Tourist travel now predominates over people as

settlers and labour. The travel is now more round trip than earlier (some states and areas discourage new settlement)¹⁹ and more frequent which means providers of passenger transport service must pay closer attention to comfort, convenience, cost and price than earlier. There is also more competition for the traveller's business than in the pre-war period in part because of higher incomes, more leisure time and an aging population.

In short, the movement of people and the movement of commodities represent two different industries, more distinct than ever before. Thus whatever regulation is applied to one in response to a particular set of problems is unlikely to be relevant to the other. This does not mean that a national transportation policy stressing efficiency or even nationalistic criteria cannot be formulated in a very general sense analogous to, for example, antitrust policy with an emphasis upon infrastructure. Whatever form such a policy may take however cannot avoid the increasing distinctiveness between the two industries.

THE ROLE OF ECONOMISTS

Another aspect in interpreting the evolution of U.S. transportation policy, and in particular the sea-change associated with the MacPherson Report in Canada and the various deregulation bills finally passed in the U.S. in the late 1970s and early 1980s, is the role of economists and economic analysis.

Economists have long had a special interest in transportation. Since much of the infrastructure has been provided and/or subsidized by governments, problems naturally arise of the proper capacity to create and how to finance it efficiently. These also include issues of pricing of publicly provided facilities, efficient user-charges, taxation, deficits in governments and so on, all problems that represent the economist's bread and butter. Transportation firms are of considerable interest as well, especially the railways with their "unusual" cost structures which early on led to investigation of imperfectly competitive markets.

The economic analysis of regulation began (in the U.S. at least) when the railroads were subjected to the *Interstate Commerce Act*, the first major U.S. private industry to be so treated. It is not therefore surprising that most of the well-known, mainstream, general economic theorists at one time or another and from a relatively early date devoted some attention

to transportation. For example, one of the founders of the marginalist “revolution” in economics of the 1870s, William Stanley Jevons, got his notion of the “margin” and emphasized the use of mathematics and statistics from his study of railroads in Australia. A. C. Pigou in his classic work *The Economics of Welfare* (1924) devoted an entire chapter to “The Special Problem of Railway Rates.” He also engaged in a vigorous debate with another mainstream economist, Frank Taussig, as to whether “charging what the traffic will bear” in railway rates was to be explained in terms of monopolistic discrimination or in terms of joint costs.²⁰ Many years later, Edward Chamberlain developed monopolistic competition theory from doing a term paper for a transportation course on the “famous” Pigou-Taussig controversy. The great Swedish economist, Knut Wicksell devoted many pages to railway pricing and rational regulation of industrial monopolies before World War I in his *Lectures on Political Economy*. Other prominent mainstream theorists such as J. M. Clark, Francis Edgeworth and Frank Knight also contributed to transportation before the early 1920s.

However, beginning in the 1920s, transportation matters elicited little further interest of important general economists. This is doubtless explained by a shift of focus to more burning questions such as the aftermath of World War I, the German hyperinflation of 1919–1923, the onslaught of the Great Depression in 1929 and World War II only 10 years later. Transport problems then retreated from the mainstream of economic analysis in favour of these more serious and vexing issues and became the subject of a small subset of economists who chose to specialize in transport matters.

This is not to say that transportation problems were not severe in the decades following World War I but so were those of most other industries and indeed the whole economy. There seemed to be nothing special about transport in this context. Various legislative Acts were passed in the U.S. that were not very carefully thought out and were generally opposed by economists, especially the *Motor Carrier Act of 1935* which seemed to be a case of wilfully misguided regulation where none was warranted, certainly by the industrial structure of trucking. However, little new analysis was presented, and the ignoring of transportation by mainstream economists meant that they were little consulted or heeded on such problems. Until the late 1950s transport economists wrote textbooks that were heavy on history, law, regulatory and carrier practices but woefully short on evaluation, analysis and critique. As has been said, “The economic theorists

believed that all the major problems in the area [of transportation] had been so well resolved that further study of them would be unrewarding.”²¹

Then in the late 1950s a series of works that applied up-to-date critical analyses to transportation matters renewed the interest of mainstream economists in transport. It signalled the rebirth of transportation as an important and meaningful branch of applied economics. Noteworthy is that this renewal was coincident with the policy changes associated with the MacPherson Report and the Message on Transportation by President Kennedy in the early 1960s. Indeed this marked the beginning of the deregulation movement and its capture by economists. Adam Smith had been rediscovered and the new microeconomics was recognized as applicable, meaningful and relevant for the over-regulated, financially strapped, differentially cartelized transportation industries.

Resources in transportation were seriously misallocated because of past regulation. Change was essential to avoid increasing economic waste and to stimulate innovative practices, new technologies and reinvigorate competition. The President’s Council of Economic Advisors began its non-stop critique of transport regulation, and President Johnson even declared before his Transportation Task Force that transportation was the nation’s number one priority! Thus began the U.S.’s long struggle to effectuate policy reform. The story will be outlined in later sections. Suffice to note here that President Johnson was side-tracked from his zeal to tackle transport policy by the civil rights movement and the escalating Indochina war. However, many economists became firmly aligned with the struggle and indeed were appointed to various government positions from which they could more directly and powerfully move forward the deregulatory agenda. The most dramatic appointment was that of Cornell economist, Alfred E. Kahn, to the head of the Civil Aeronautics Board (CAB). Having earlier declared that “applied microeconomics is the exciting new frontier of public policy,”²² he went on to promise application of rigorous efficiency criteria to all regulatory matters and ultimately eliminate economic regulation entirely, and with it the CAB itself. He was hugely successful.

By 1980, the regulatory reform movement had achieved substantial success with both air cargo and passenger transportation rid of “artificial” economic constraints two years earlier. The firms were given (some with reluctance) free entry and exit and pricing freedom subject only to the antitrust laws

and safety regulations. In 1980 the railroads were provided a higher degree of pricing freedom and easier conditions of exit²³ and motor carriers got freedom of entry as well as exit, thereby eroding the scarcity value of previously granted operating certificates, along with a large dose of pricing freedom. Household goods carriers were similarly deregulated in 1980 and interstate bus operations in 1982, as a kind of afterthought so typical of the policy approach to passenger transportation.

It is interesting to note that the economic arguments leading to transportation deregulation and their subsequent success in both Canada and the U.S. stimulated the worldwide swing toward greater reliance on market forces, deregulation and privatization which ultimately led to the utter disparagement of the kind of socialist and communist economic planning replete with state-set prices and monopoly under petty tyrants and unpleasant dictators found in Eastern Europe and much of the rest of the world. It also presaged the end of the war between communism and capitalism since the former could neither produce the goods and services nor provide freedom or much security outside of prison. From Khrushchev's empty threat that "we will bury you" to Gorbachev's "we will deny you an enemy" implies a massive leap in visions of appropriate economic institutions. Much of the rethinking of the microeconomics of regulation and the emphasis upon market-determined prices and efficiency as the desired over-riding criterion of public policy towards transportation that took place in North America from 1960 on, and the successful implementation of policy change, paved the way for the worldwide reemphasis upon market forces. This was indeed "the economist's hour."

2. THE PERIOD BEFORE 1930

The two main aspects of federal government transportation policy between 1930 and 1991 were the development of the transport infrastructure and the terms and conditions for the use of such infrastructure, namely regulatory policy of both the economic and safety variety.

INFRASTRUCTURE POLICY

Highways

Although early involved in building the Cumberland Road, completed in 1844, the federal government withdrew from further highway construction for almost 50 years on the grounds that such outlays were unconstitutional

and, in any event the states' requests for highway funding were many times the size of the federal government budgets at the time. Such roads as were built, sometimes stone overlaid with gravel but usually dirt piled upon logs (corduroy roads), plank roads or simply scraped and widened dirt trails, were mostly financed by state and local communities and operated as turnpikes. Sometimes private companies built roads but many of these reverted to state or local agencies when profit prospects dwindled or disappeared. The railway era especially after 1850 reduced the need and incentive to produce highways for the then existing horse-drawn vehicles. The railways had both a better right of way and the requisite vehicles and horseless motive power, to boot. Furthermore the railway right of way was privately financed although heavily supported by federal land grants. Further development of road transportation clearly awaited a new vehicle and especially a new method of propulsion which became available only in the early 1890s. Yet the highways of the U.S. were probably in worse condition through the early 1900s than they had been 40 years earlier when they were generally described as atrocious.

The reason for the demise of the road system, or rather road segments and patches, was the loss of traffic, both passenger and freight, to the railroads which, from mid-century on, dominated transport in the country. Railroads are believed to have provided some 95 percent of intercity passenger trips by the 1890s. This destroyed the turnpike companies which let the roads fall into a state of almost complete disrepair. Ultimately the responsibility for these roads reverted to local government which generally was ill-equipped and too financially strapped even to maintain let alone to expand them. The position of county road overseer degenerated into a mere sinecure. The highway "system," which had shown some signs of acquiring a real inter-state character, became fragmented and run down. This was no great threat to national unity nor economic progress at the time because of the vigorous and expanding railroad system; it was the more localized and intrastate transport needs that progress had temporarily by-passed.

The inadequate and deteriorating road system became increasingly intolerable after 1890. The demand for improved and extended overland right of way soared for three main reasons: the bicycle craze, the initiation of free rural mail delivery, and the development of the automobile.

The "good roads movement," as it came to be called, sought to eliminate the disparities between rural and small-town living on the one hand and urban areas well connected with the national economy on the other. Since

railways by their very nature could not be as ubiquitous as roads, this seemed like a good place to start. By 1890 the bicycle manufacturers had begun their campaign for road improvement. Cyclists obviously wanted passable roads, and road improvement was believed to be important for national economic growth. Surprisingly the railroads strongly supported the bicycle crusade believing that the national economy would benefit, thereby increasing their business, and that road transport could never be anything but ancillary to the railroads.

Free rural mail delivery, after a false start, resumed in earnest in 1896 to provide service to villages with populations of less than 10,000. By 1900 over 4,000 routes had been established. Since the Post Office Department had ruled that service was contingent upon the availability of adequate roads, residents began clamouring for them more and more vigorously. Realizing the importance of this service for national unity as well as the obvious political, social and economic advantages of a more ubiquitous and diversified system of communications, the federal government proposed to make financial aid available to state and local governments. It was not however until 1916 that federal legislation was passed.

In the meantime the third factor in the rise of the public demand for good roads, the automobile, was beginning to make its appearance. Although only 8,000 vehicles were registered by 1900, the potential demand was already obvious. The big need was decent roads on which to drive the new machines. Thus the automobile manufacturers and owners' associations added their voices to the good roads movement.

As a result of these efforts, state aid began to be offered for highway construction. The state of New Jersey in 1891 argued that "public roads within a township were constructed for the convenience of the citizens of the counties in which they are located, and of the entire State as well of said townships."²⁴ The law also provided that the cost of road construction would be split among the adjacent property owners (10 percent), the state (33 1/3 percent) and the county for the remainder. Other states followed suit so that by 1917 all 48 had some form of highway aid program.

To provide for interstate needs, the federal government needed to become involved. Earlier it had organized the Office of Public Roads Inquiry to undertake research in road construction methods and to publish the results.

However, when it came to authorizing public funds for highways, considerable opposition arose which was not successfully repulsed until 1916 when President Woodrow Wilson signed *An Act to provide that the United States shall aid the States in the Construction of rural post roads, and for other purposes* or, more simply, the *Federal Aid Road Act*.²⁵ The rationale for federal assistance was the more effective performance of government functions and although not clearly stated, the promotion of interstate commerce.

It is perhaps not coincidental that by 1916 more than 2.5 million vehicles were registered in the country — a sharp growth over the 8,000 in 1900. Even without federal aid the automotive revolution had begun with a vengeance. With such, on top of state and local financing, the highway boom took off. To secure federal aid each state was required to assent to the provisions of the Act and create a state highway department capable of supervising construction and maintaining the designated highways. The aid was limited to post roads and in an amount not to exceed half the cost. The 1916 legislation was supplemented in 1921 by the *Federal Highway Act* that promoted an interconnected national highway system by apportioning funds to each state on a formula basis, the exact amount however depending on Congressional appropriations. Partly through these stimuli as well as the generally prosperous conditions of the 1920s, the mileage of surfaced highways grew from 154,000 in 1904 to almost one million by 1934.

Railways

The story of the railroads is more familiar and will be treated briefly. Passenger miles peaked in 1920 and fell through 1933, the depths of the Great Depression. Even before the beginning of this severe contraction, traffic had declined over 50 percent by 1929 indicating that this was not merely a cyclical phenomenon. Indeed it continued, except for the World War II years, until the railroads succeeded in divesting themselves of the passenger business and most of the directly related equipment, much of it antiquated, to the new government-formed National Railroad Passenger Corporation (Amtrak) in 1971.

The development of the railway infrastructure was significantly stimulated by federal aid mostly in the form of land grants. Although the concept of land grants was not new, having been used on a modest scale for canal

development and navigation improvements on rivers and selected highways earlier, it was employed on a vastly expanded scale in the case of railways after 1850. Ultimately over 179 million acres were given for railway construction, an area larger than Texas. The purpose of the land donations, beyond that needed for the right of way, was to provide an asset basis against which construction costs could be borrowed or as backing for loans and stock sales. The government retained alternate sections of land in a strip on each side of the rail line to sell later in compensation for the land granted to the railroads at inflated values caused by the economic growth expected to be stimulated by the new transport access. In addition, the companies were not only required to build and operate the railroad for which the grants were made but, in many cases, to move government property and troops over the line free of charge or at reduced rates. Such a subsidy, now estimated to have been totally repaid through the value of reduced rates for government traffic over the years, stimulated railway construction not only in the eastern section of the country already fairly well served by rail, but especially in the west. In fact it opened up the west much more rapidly than would otherwise have occurred. It was part of U.S. "manifest destiny" not unlike the "Canadian dream."

The rapid growth of rail capacity was mainly justified by military needs to secure the west coast from British intrusion as well as to develop the country more fully and exploit not only the land mass thereby made accessible but also the rapidly evolving rail technology. The improved technology plus the sharp rise in traffic reduced average revenues per ton-mile from about eight cents in 1848 to less than two cents in 1870. Costs fell even more in response to efficiency and volume gains. Rail mileage burgeoned from roughly 9,000 miles in 1850 to 53,000 in 1870. It tripled by 1890 and ultimately reached about 253,000 miles, its peak, in 1920. This rapid expansion combined with increased capacity of the track as well as continued improved efficiency of the rolling stock and locomotives, led to such enormous excess capacity that even the rapid growth of the U.S. Gross National Product at the time could not utilize it profitably.

A wave of mergers and consolidations accompanied the rapid expansion of capacity with the resultant development of monopolistic practices — known in railroad parlance as pools, traffic associations, or rate agreements — to maintain not only a stable *level* of rates and fares but a *pattern* of discrimination, euphemistically called the rate structure. Castigated by economists

as “value-of-service pricing,” such activities caused public outrage and pressures for regulation as early as 1869 in Illinois. This was to culminate in federal regulation in 1887, discussed below.

Other Transportation Infrastructure

The federal government has only reluctantly taken a supportive role to assist new transport technology whenever its implications seem to stretch beyond the reach of private development even with the help of state and local governments. There is little tendency to examine how the new technology can be made in some sense to “fit in” with what exists; nor has there been much success at evolving an overall “plan.” In short, the federal role has remained ad hoc with respect to technology and specific investments emanating from that technology be it highway, rail or air. Up to 1930 at least, the federal government’s approach was of modal separation with no major attempt at intermodal coordination. Furthermore the federal role in transport was tepid and reluctant, including only “projects of truly national importance which absolutely required federal intervention to succeed.”²⁶ Each such intervention involved the private sector in the major role of construction, operation, and/or principal risk taker or initiator. None was subject to much economic analysis, most were reactive and all were rooted in optimism, patriotism and nationalism.

In a *laissez-faire* environment, with limited, constitutional government, this was a fitting approach. However, the Great Depression changed all this.

Thus, the pattern emerged whereby initiatives for new technology, arrangements or institutions began in the private sector. If more resources were needed than the private efforts could muster, the next level of recourse was local government, then state and only as a last resort would federal government help be sought and then only if interstate or national interests were clearly involved. This is quite evident in the cases of federal support for highway and railway infrastructure. A similar pattern emerges where certain aspects of private activity are deemed antisocial or detrimental to the economy in some way. The very creation of the Interstate Commerce Commission provides a good example. In all the major transport issues the ultimate involvement of the federal government took many years. In cases of investment or finance, the assistance was piecemeal and specifically circumscribed. No blank cheques were written. Few of the infrastructure

investments involved economic or efficiency criteria. Certainly there were attempts to ensure some form of repayment, such as tolls for the Cumberland Road and the alternative parcel of land holdings in the land grant schemes for the railways as well as some other forms of *quid pro quo*. In this, a semblance of "sound" business judgement was present and was used to justify the actions taken. But no overall criteria of investment-type aid included economic calculation. The criteria were mostly defence, some notion of improvements to interstate commerce, opening up the west, manifest destiny and other aspects of nationalistic sentiments. All of these were also quite vague.

This response to pressures from the bottom up does not mean the federal government could not take the initiative when circumstances warranted. The seizure of the railroads during World War I is an example. But by and large the federal government pursued a pretty rigorous, *laissez-faire* attitude when it came to transportation of people and goods and even infrastructure until the situation virtually cried out for involvement at the highest administrative level.

This contrasts sharply with policy initiatives taken in subsequent decades at least through the 1970s. Only during the 1980s and especially in the latest transportation policy pronouncements from Washington do we see an attempt to return to a more bottom-up approach. Largely this has to do with the ideological revival of a thoroughgoing willingness to rely on market forces. It is however closely linked to the budget deficits that have beset the U.S. (and Canadian) federal governments. The return of transport policy more and more to the states and cities reflects a combination of ideology and dearth of financial resources. Policy towards transportation thus appears to have come full cycle. Nor is there much more formal economic analysis regarding transportation infrastructure than in the pre-1930 period. It is not yet recognized in U.S. federal budgeting that expenditures on roads, ports, canals or other physical assets should be treated as investments. There is still no distinction between current and capital outlays in the U.S. accounts. Given the advances made in economic appraisal of public investments, there is much less excuse for failure to do a better job of pre-feasibility analysis than in the past and to continue to rely on vague criteria concerning the national interest, prestige and the usual standby, national defence. Nor is there much merit to permitting the infrastructure to deteriorate or refrain from adding to it when warranted because of fears of raising the deficit.

Another factor forcing the federal government to respond to economic forces in general and transportation in particular was the growing interrelatedness of the economic system. Transport investment was both cause and consequence of this. And with it came the phenomenon of big business which required higher levels of government and bigger size thereof to cope with the problems. Certainly as interstate commerce grew faster than total economic activity, neither local nor state governments had the appropriate jurisdictions to deal legally with ever bigger, interstate companies. The federal role had to increase. The Great Depression simply made it more imperative and urgent. Prior to this it called for different forms of intervention which were ill-understood at the time.

REGULATORY POLICIES

The latter part of the 19th century and at least until World War I have been described as the “high tide of *laissez-faire*.” Keynes exclaimed, “What an extraordinary episode in the economic progress of man that age was which came to an end in August 1914!”²⁷ Of course it was not such a wonderful era for most of mankind nor even for those living in the fortunate part of the world to which Keynes was referring. Nevertheless the prospects that Marshall held out for “opening up to all the material means of a refined and noble life” were not believed to be far away in the glow of Victorian optimism. Capitalism seemed to be on a roll that even the churlish thunderings of Karl Marx could not deflect. U.S. economic growth from 1850 to the Great Depression was phenomenal. Part of this was no doubt due to the railroads. True, there were those who subsequently questioned whether the railroads really contributed so much. In fact some were unkind enough to suggest that the incredibly rapid growth of railways used up excessive amounts of resources, such as iron and steel, that prices were raised excessively and scarcities occurred to the detriment of development of other manufacturing industries. Government support of accelerated growth of the rail network thereby may have slowed down development of the economy elsewhere and perhaps overall. But these caveats were to come later.

At the time, railroads were viewed as the engine of economic growth to say nothing of national cohesion, manifest destiny and the end of poverty. Part of this was doubtless true. At the same time however the railroads had by the 1860s overtaken in size and even influence, most other institutions in U.S. society. They employed many times the number of people than did the

federal government. The value of their assets was far greater than those of the largest industrial corporations. It has been argued that their social influence in the late 19th century was akin to that of television in the late 20th century or even Catholicism in medieval Europe.²⁸ It is true that the railroads created time zones, emphasized punctuality, seemed to be the essence of economies of scale and hence “natural monopolies,” dominated many a state legislature, browbeat and swindled local communities and otherwise rode roughshod over much of American society, economy and polity for several decades. This posed an enormous challenge to free markets to say nothing of individual freedom and political democracy. Big business first reared its ugly head in the form of a railroad, which the federal government had even assisted in creating!

Partly due to its economic characteristics, namely the large proportion of its costs that are joint, fixed and/or otherwise indivisible or non-allocatable on cost-occasioned criteria, a railroad must charge a price, rate or fare for particular traffic that is differentially in excess of marginal cost. In its most sophisticated and socially acceptable sense this is referred to as “Ramsey pricing” or maximization of profits subject to a revenue constraint which is now not only permitted but legally required by the *Staggers Rail Act of 1980*. In its cruder form before 1930 it was viewed as flagrant discrimination which it often was. One form of it especially annoyed farmers, namely when railroads charged more for a short-haul movement of the same commodity than for a longer haul over the same line and in the same direction! Such a violation of the view that prices should bear some, however vague, relation to costs aroused passions. It seemed manifestly unfair. This kind of pricing combined with secret rebates to favoured shippers, frequent and apparently whimsical rate changes, chronic rate instability, general railroad arrogance and arrangements among railroads not to compete on price or service as well as to pool traffic — much like OPEC attempts to assign oil production quotas at present — led to considerable agitation on the part of small shippers mostly of farm products especially in the mid-west. They succeeded in convincing several states, notably Illinois, to pass what came to be called Granger Laws which set permissible freight rates by a regulatory commission or a legislature, prohibited long-short haul discrimination as well as free passes to public officials and in other ways sought to restrict certain dubious practices that had emerged in the provision of rail transport. Although later repealed, most of these laws were upheld on the constitutional grounds that rail transport was “affected with the public interest” and that railways

were in fact “common carriers.” The latter is a concept dating far back in British common law and creates certain obligations upon those holding themselves out to provide the public with some specific service such as inns, transport providers and others. It implies among other things that prices will be published and not changed without prior notice, that all who request the service will be provided without discrimination at such prices, that common carriers assume certain liabilities for loss and damage and other transportation-related matters.²⁹

Federal action was finally required when the courts held that a state could not control rates on interstate traffic. Since most of the traffic by this time (1887) was interstate, the need for federal legislation was apparent. Finally, and after many years of effort, the *Act to Regulate Commerce* was passed in 1887 by Congress. It stressed that the primary evil to be remedied was freight rate discrimination although it applied to passengers as well.

There remains some controversy concerning the “real” reason the railroads became the first major private business to be subjected to federal commission-type regulation. The conventional wisdom has it that the companies so abused their excessive amount of economic power that the government was forced to retaliate by special regulation, even of otherwise sacrosanct private property, to eliminate abuses and provide suitable punishment as a preventive. It has more recently been argued that the railroads themselves invited regulation to implement a more effective cartel since their pooling arrangements frequently broke down and “cut-throat” competition erupted.

The truth seems to be that railroad arrogance and behaviour were the ultimate reason for the introduction of regulation. People had been so aggravated along with obvious abuses, that for the previous 20 years some 150 separate bills had been introduced in Congress. Finally when the Wabash case invalidated state regulation of interstate commerce, some kind of federal control seemed likely to pass. At this point the railroads recognized that “we must make up our minds to it” as one railway president put it. After that, the attention of the companies focussed on making the regulation as palatable and innocuous as possible. At least something better than the tough state regulations seemed preferable.

And so the Act was passed. President Cleveland signed the bill which declared railroads to be common carriers; created a five-man commission; outlawed pools and rate discrimination; required that all rates and fares be

just, reasonable and published; authorized the Commission to investigate any interstate railway; empowered it to compel witnesses to testify and to secure relevant documents; and required railroads to submit regular reports and adopt a uniform accounting system. At the end of March 1887, with five Commissioners and 11 staff members, the Interstate Commerce Commission set out to eliminate abuses of economic power by the largest industry in the country, at the time employing about 800,000 workers and having over 100,000 miles of track and abundant other capital assets. Hope ran high. One newspaper proclaimed that “the *Interstate Commerce Act* is working wonders and the rail magnates are trembling,” after several early decisions. Alas this was not to be the case until over 20 years later.

As the Commission sought to enforce its rulings and to behave as it correctly believed Congress had desired, the Supreme Court systematically emasculated all of the authority the ICC thought it had in a series of rulings extended over the next 15 years. By 1903 the Commission lamented that all it could do was “investigate and report and such orders as it can make have no binding effect.”

This was also the era of the so-called “first merger movement” which radically transformed the structure of American industry. The merger mania, it should be noted, took place within a short period of time following the passage of the first federal antitrust law, the *Sherman Act*, in 1890. The attempts to restrain the development of monopoly power in the first place fared no better than the attempts to restrain its abuse once “naturally” acquired as was presumed in the case of railroads! The U.S. government was thus called upon to seek to make the 1887 transport act effective and at the same time create a climate wherein most other industries would be restrained from cartelization or at least from more industrial concentration than had already occurred between 1897 and 1904, the period of the first merger movement.

Antitrust was more difficult to make effective because the courts were unwilling or unable to unscramble the corporate eggs once blended, and fine legal distinctions began to be made between “good” and “bad” monopolists. The antitrust laws with no regulatory commission or agency at this time (the Federal Trade Commission was created in 1914) were to be enforced by the Department of Justice through the courts, something that commission-type regulation was seeking at least partially to avoid but which the ICC to this point had not yet accomplished.

The Interstate Commerce Commission however fared better than did anti-trust activities. By 1910, through a series of new enactments, Congress restored to the Commission all the original powers it thought it had previously given. At the same time the courts limited their right of judicial review of ICC findings and judgements. This change occurred because of the immense support by the activist President Theodore Roosevelt of railway regulation and the ICC in particular, as well as the growing public opposition to the railroads and the other "malefactors of great wealth." In any event, the ICC had a second chance. Unfortunately the Commission squandered its opportunity and with its now considerable strength it essentially preserved the *status quo* with regard to the rate structure and collective rate-making by the carriers themselves in rate bureaus. It also denied a series of general rate increase requests by the railroads or sanctioned lower increases than in retrospect were needed.

The exigencies of World War I and confusion and division within the ICC induced President Wilson to exert federal control over the whole railroad system in order better to co-ordinate traffic, among other things. The ICC was not believed capable of doing the job. The federal Railroad Administration was set up in 1917 and by its general Order No. 1 the Director General of Railroads required the pooling of all equipment and facilities, ordered new routings regardless of previous arrangements, denied shippers the right to route and sought to operate all railroads as a single *national* system.

At war's end, the shippers clamoured for the return of the railroads to private operation and ICC regulation probably because the Commission had resisted rate increases despite an acknowledged need that was acquiesced to under government controls. Others wanted nationalization and fashioned plans that received some attention. However, *laissez-faire* principles were so firmly embedded that not only were the plans rejected outright but attempts to extend the period of federal control were ignored. Instead Congress passed the *Transportation Act of 1920*. This broadened ICC authority, legalized railroad combinations subject to an overall "plan" that the Commission was to devise, enshrined the rule of rate-making (fair return on a fair value) into law and gave the Commission control over security issues and service.

In general the idea was to rationalize, definitely *not* nationalize, the system and abandon the emphasis upon competition among railroads and between them and the rapidly emerging truck and bus industries. Inter- and intra-modal

coordination was the new fetish probably inspired by the credible job done by the Railroad Administration during the war and the equally new fetish from which coordination derived, namely "scientific management." Even Joseph Eastman, soon to become an ICC commissioner and destined to be its greatest, approved of both extension and nationalization. But it was not to be. The *Transportation Act of 1920* hoped that private ownership and operation under the "inspired" leadership of the ICC would yield most of the advantages of nationalization without the disadvantages.

This was not to be either. The Commission was reluctant to exert any leadership and refused to formulate a formal plan for railroad reorganization ostensibly to eliminate the so-called weak-strong road problem. The ICC even begged to be relieved of the task of bothering to concoct any plan. The Commission did not "free the industry . . . of the accumulated burdens of past ills . . . [nor] establish principles and practices for this regulated industry beyond and above those recognized in the general competitive field."³⁰ Nor did the Commission urge the railroads to reduce bonded indebtedness or to consolidate into fewer more economically viable systems. This meant that when the Great Depression began, the industry suffered far more from reduced business than would otherwise have been the case, and the Commission, regardless of its legal authority or of any infusion of leadership zeal not hitherto apparent, was totally unable to help out under such overwhelming economic forces.

Conclusion

On the eve of the Great Depression, the U.S. possessed a substantial rail system that was heavily concentrated in the east and sparse but adequate in the west and much in between. The mileage had already begun to shrink to eliminate much redundancy in particular parts due mostly to the frenzied over-building especially between 1893 and 1920. The beginnings of an increasingly ubiquitous highway system had been well established which hastened the rail redundancy. Passengers had growing options for intercity movement in many parts of the country but especially the eastern section. The rail monopoly had been effectively broken and was never to be regained in either passenger or freight business although this was not recognized at the time.

The railroads, for the most part, supported the alternative highway mode on the assumption, often stated, that the new service using highways could

never be anything but ancillary or “feeders” to rail connections for ongoing journeys or goods shipments. It would also relieve railroads of the short trip or haul which was the less profitable phase of the business anyway. By so doing, trucks and buses and even private automobiles should be encouraged for reasons of profitability as well. Further highway construction and the industries producing the vehicles would add to rail traffic as the raw materials and finished products would inevitably be shipped by rail, except for the minor exceptions in a few parts of the nation where waterways might be used. The stimulus to overall goods production and incomes generated by the new mode and its infrastructure would also raise the growth rate of GNP which would further stimulate railroad traffic.

Thus instead of opposing the new mode, the railroads supported it, at least in a broad, general sense. They even invested somewhat in bus transport to serve light-density areas without branch lines or to permit branch line abandonment in many cases. The Pennsylvania railroad helped create the Greyhound network, and other railroads owned part of the National Trailways bus system. By 1941, rail interests are reported to have owned 1,759 buses serving 44,700 miles of route.³¹ Railroads invested very little in the trucking industry and viewed it as even less of a threat. It is hard to see this as especially short-sighted from the vantage point of the late 1920s. Air travel was certainly no rival yet and the highways, pipelines and waterways had not yet made such inroads as could topple the mighty railroads from their pre-eminence. Besides traffic was growing rapidly during the Roaring Twenties so there was plenty to go around even if one's relative share was shrinking. On the other hand, rail short-sightedness was monumental by virtue of hindsight. Had they taken over the short-haul trucking or bus businesses to a far greater extent, which was then feasible, they would have received grandfather rights in the less likely event that highway service would have been regulated. In either case, the extent of modal integration might have been far greater than became the case in the U.S.

Failure to enter more fully into bus and truck transport also supported development of independent truck and bus operations which heightened the emphasis upon modal separation despite all the stress that came to be put on coordination and integration. This is all speculative, of course, but is one of the “if only” myths as noted in “Icons and Albatrosses.”³² It should also be remarked that any major entry of railroads into truck or bus transport might well have led to several massive multimodal oligopolies whose

contributions to efficient resource allocation by mode might well have thwarted developments in truck and bus transport. Intra-firm efficiency might have left much to be desired by such large multimodal enterprises whose (largely railroad) experience and sympathies, to say nothing of relative investment commitments, might have had strong negative effects.

THE POLICY OF MODAL SEPARATION

The reticence of rail companies to enter into truck or bus operations extensively or to acquire large numbers of existing enterprises in highway transportation combined with the national fear of excessive railway concentration of economic power, already enormous before 1887, led to legal and regulatory restrictions against multimodal companies or what later came to be called integrated transportation firms. As early as 1912, long before highway modes were anything more than a distant possibility, the *Panama Canal Act* was passed forbidding railroads from operating or having any interest in water carriers using the Panama Canal without specific ICC approval. The fear was that such ownership would restrict the development of intercoastal water traffic via the Panama Canal to protect the rail monopoly over the coast-to-coast traffic by land.

Similar beliefs arose during the 1930s with respect to rail acquisition of trucks and buses. The *Motor Carrier Act of 1935* sought to preserve the "inherent advantages" of motor carriers, a corollary of which required that any motor carrier acquired by a railroad must be used "to public advantage in its operations." This was promptly interpreted by the Commission as obligating the railroad to operate the motor carrier entirely as an adjunct to its train service.³³ Congressional intent was to prevent the acquisition of control of motor carriers so that it not "get into the hands of other competing forms of transportation who might use the control as a means to strangle, curtail or hinder progress in highway transportation for the benefit of other competing transportation."³⁴ There was, in short, concern that the railroads would acquire other modes solely to protect their own much greater investments in rail facilities. It was feared they would stifle developments of other modes except as small-scale contributors to rail operations and would perpetuate rail monopoly power and profits. There was little recognition that a multimodal enterprise could contribute to overall efficiency or that in many respects the modes were complementary rather than competitive. Thus the tradition and even policy of modal separation in the U.S. persisted until the early 1980s, well after the deregulation movement had taken place.

This is all the more surprising since enthusiastic support for the integrated transport firm selling transportation by rail, truck, bus, water or any other mode presumably using the least-cost combination for any shipment or travel, began early in publications by H. G. Moulton and The Brookings Institution.³⁵ These were reiterated later. For example, the Doyle Report (1961) flatly stated that “there should be no prohibition in the law against the formation of . . . a transportation company . . . a carrier of any mode [should be allowed] to become part of such a transportation company.”³⁶ In 1979 the National Transportation Policy Study Commission recommended the elimination of “Federal impediments to common ownership and inter-modal coordination and cooperation,” and the promotion of “effective joint rates and through service within and among modes.”³⁷

Despite removal of most formal restrictions against intermodal operations and ownership, there are few examples in the U.S. today of large multimodal enterprises and nothing analogous to each of Canada’s two major railroads owning and operating large national trucking companies. Indeed, in Canada, multimodalism has a long history with the CPR chartering ships to carry traffic on the Pacific Ocean in 1886, inaugurating passenger service there in 1891, passenger and cargo services on the Atlantic in 1905, and later truck and airline services as well. The MacPherson Commission found rail ownership and operation of trucking services in Canada consistent with efficiency and saw “no reason to limit the entrance of railway companies into any other mode of transport. The experience of other countries with such restrictions does not encourage us to recommend it to Canada.”³⁸ So much for the U.S. position on modal separation!

The more or less naïve U.S. position on modal separation, is believed to have retarded the growth of piggyback (TOFC [Trailer on Flat Car] and COFC [Container on Flat Car]) operations in the U.S. as compared with Canada and otherwise to have impeded efficiency achievable from greater modal integration. On the other hand, at least at the beginning, the belief that the railroads might have misused or wasted, from the social point of view, any large-scale investments in truck or bus operations was not so far-fetched as it seems.

Since rail intercity passenger traffic had not accounted for much over 10 percent of total rail revenues during the 1920s and an even smaller proportion of profits, in the face of a sharp downturn in number of passengers and

passenger-miles, it is difficult to see why railways continued to stress investments in and general support of this segment of the business. Certainly, the ICC did little to require continuance of unprofitable service even after being given what has been referred to as "sweeping" powers over abandonments. Indeed, the Commission "Guided primarily by financial considerations . . . generally approved abandonments."³⁹ Maybe there was something to the "image" in the sense that despite lower profitability, perhaps some railroads felt that the prestige of grandiose terminals, the luxury side of passenger travel that provided association of the railroad executives with the rich and the famous in America was worth the expense. At least it suggests that railroads would have been unlikely to stress development in truck or bus subsidiaries had they chosen to move vigorously into these areas during the 1920s when the opportunity presented itself.

For the latter 1920s the railways were still profitable, possessed much monopoly power in numerous freight and passenger markets and were still the giants of U.S. enterprise. After return to private ownership and the other features of the *Transportation Act of 1920*, they were expected to remain so. The Act further enhanced the powers of the ICC to prevent "cut-throat" competition, sustained collective pricing agreements (i.e. rate bureaus) contrary to the antitrust laws, facilitated abandonment of unprofitable service and otherwise benefitted the railways under the new so-called "positive" approach to regulation. The latter involved helping resolve the "weak-strong line problem," stimulating mergers between profitable and less profitable or losing railroads and, in fact, aiding in the accomplishment of the new rule of rate-making which sought to ensure to each carrier, as a kind of *quid pro quo* for being subject to regulatory "restraint" and more or less ignoring the benefits of past public largesse, a "fair return on a fair value" of the property devoted to provision of rail transportation service. This, of course, meant enormous profits or yields on assets of dubious value for an aggressive and lucky few in the likely event that the Commission failed to develop a sensible plan of railroad consolidation.

All of this was speculation. In less than a year of the close of the decade of the 1920s, the economic and, indeed, social and political world of the U.S. and almost everywhere else was suddenly and irrevocably changed beyond all belief. No one foresaw the Great Depression, least of all economists or governments.

3. THE GREAT DEPRESSION

The 30 percent contraction of real GNP from 1929 to the trough in 1932-33 was the most severe to beset the U.S. economy before or since. It was also the longest, with the level of economic activity remaining persistently below that achieved in 1929 for 10 years and longer than that in terms of GNP per capita. The reasons for the severity and duration are not yet altogether clear. From the U.S. point of view, however, what would doubtless have been a more or less routine contraction in 1929 was turned into a rout probably because of three major macroeconomic policy errors that it is hoped will never recur. In the first place Congress passed the Smoot-Hawley protective tariff Act in 1930, which understandably led to sharp retaliation from the U.S. trading partners. Such "beggar-thy-neighbour" policies succeeded in shriveling up much international trade including, of course, U.S. exports. Second, The Federal Reserve System allowed the money supply to shrink by about 30 percent which kept real interest rates too high and promoted financial panic. Third, the income tax rate was increased in 1932 to balance the budget. This reduced disposable incomes and contributed to the decline in personal consumption expenditures, the single biggest component of GNP.

Whatever the causes of the excessive severity of the depression beyond the probable "normal" cyclicity of market systems, the results for the railways were devastating. Passenger traffic dropped by almost 50 percent between 1929 and 1933 while passenger revenues collapsed by about two thirds. Freight performance was almost as bad. Despite declining operating expenses, net income fell from a positive \$977 million for all roads in 1929 to a negative \$122 million in 1932. It remained extremely low for most of the 1930s and only exceeded the 1929 level in 1942 during the wartime boom. Bankruptcies rose during the depression despite generous government aid and in 1939 included almost one third of the total rail mileage. Passenger deficits began in earnest and were the main reason for the overall deficits and minuscule net revenues. Starting with modest negative earnings in 1930, the passenger deficit rose to the \$250 million range during the years 1938 through 1941 and constituted some 30 percent of net revenues from freight traffic on average until 1941.

However, while rail executives believed that the dismal passenger performance through 1933 was largely due to the depression, total intercity passenger travel had only decreased about 20 percent since 1929 compared

with 50 percent for rails. In fact railroads suffered more than any other mode in extent of decline of business,⁴⁰ as was pointed out by Joseph B. Eastman, the Federal Coordinator, a post created by the *Emergency Transportation Act of 1933*.⁴¹ His Report also asserted that, important as the depression was, "Flight of passenger traffic from the railways is due to failure to keep pace with modern methods of marketing, servicing, pricing and selling."⁴² The railroads were lacking in "personal helpfulness," depots and terminals had been designed mainly for "operating convenience" rather than "passenger convenience," marketing policy among the railroads was competitive rather than coordinated. The entire attitude of rail personnel from management down was obsolete and needed radical change if the "great opportunities for the reestablishment of railway passenger service upon a profitable basis" were to be seized. Such opportunities were believed to exist because the growth of automobile traffic "indicates the volume possibilities which await a still more attractive and economical mode of travel."⁴³

Many other suggestions for reform were made in the Report of the then powerful Federal Coordinator. However, as has been pointed out, "rail executives resorted instead to 'flash' over substantive reform and hoped that their problems would go away."⁴⁴ The new strategy was called "streamlining" and involved development of extraordinary passenger trains such as the Union Pacific's M-10,000, and the Burlington Road's *Zephyr* which captured the public's imagination at the Century of Progress Exposition in Chicago in the summer of 1934. Movies were made of the new trains and a new streamline craze became for a while attached to numerous types of product.⁴⁵ For a short time, the downward trend in rail traffic was arrested. Both the number of passengers and especially passenger-miles rose from the depths of 1933 by 15 and 52 percent respectively through 1937. However, traffic dropped off again until the brief wartime revival then resumed a continuous decline through 1970 before the creation of Amtrak. "While the success of the streamliners provided reason for optimism, all was not well in the railroad passenger business."⁴⁶ Indeed, passenger deficits stayed high through 1941 despite the traffic gain and resumed their upward spurt after 1945. More needed to be done along the lines recommended by the *Passenger Traffic Report*. By the time the railroads realized this, it was already too late.

On the other hand, there were more positive aspects for rail rivals largely because, in seeking to moderate the depression, the federal government embarked on massive, expenditures many of which went for transport

infrastructure in highways and airways. In fact, the quality and quantity of the highway system were considerably improved by World War II over what had been its state at the start of the decade. Air transport also received a big boost not only from direct subsidies but from development of the airways and airport systems as well during the depression. Thus two of the railroads' leading rivals received right-of-way expansions that made their future inroads upon rail traffic far more potentially severe and certainly much earlier than would otherwise have been the case. Buses and private cars siphoned off rail passengers especially for short-haul traffic while airlines were preparing themselves, also at public expense, for attacking the long-haul passenger business. Truck traffic was to capture increasing amounts of the freight business but more important, the higher rated and valued traffic was rendered more susceptible to such "theft" by the railroads' misapplication of value-of-service pricing. That is, freight rates and some passenger fares were kept further above marginal costs than demand elasticities warranted especially when the latter increased as inter-carrier competition grew.

Thus the general rail advantage of lower marginal or variable costs was often dissipated by excessive and not necessarily profit-maximizing mark-ups over a specific unit cost, itself over-estimated by inclusion of arbitrary amounts not related to avoidable costs but rather arbitrary accounting allocations. In short, the railway pricing system at the time was crude to say the least — a by-product of monopoly power and general lack of competition and marketing savvy as well as inadequate knowledge of specific avoidable costs. A large number of rates, perhaps as many as 30 percent were, even in a later and more enlightened era, found to be below the ICC version of actual variable costs! Many more were excessively high in terms of maximizing the net revenues or net profits. The rate structure made little economic or financial sense.

The arrival of new competitors on the scene, with not only different costs but generally superior quality, meant that without significant reductions in particularly high rates, far more traffic would be lost than was either necessary or, from a social point of view, efficient. Below cost rates needed to be raised. Clearly the "preservation of inherent advantages" of each mode, required by the 1940 Act, could only take place under a regime where specific rates and fares reflected correctly the underlying marginal costs and the specific demand elasticities. Some mark-ups over costs were needed to finance the many joint, fixed or otherwise indivisible costs so characteristic

of railway operations. This was, of course, too much to expect during the era of rail monopoly when the rate and fare structure was established. The incentive to create an economically rational structure did not exist. Nor, it should be noted, did the techniques. The period of deep depression followed by the war-induced traffic boom failed to provide incentives to price rail services correctly either even if efficiency had been the goal. Only after these overwhelming events could rational traffic allocation via efficient pricing have a chance to succeed.

The attempt to achieve allocative efficiency by *extending* the scope of regulation to all modes and providing the ICC and CAB with far reaching powers was not shown to be fruitless and enormously costly until after the late 1940s. Thus, in a major sense, the Great Depression provided fertile soil for the “grand experiment” or what I have called elsewhere “the grand Benthamite design”⁴⁷ involving detailed regulation of all modes and most aspects of transportation under private ownership. The failure of this experiment engendered the reformist zeal of the late 1950s. The depression itself represented an enormous failure of the market system, even if its severity and duration had other contributing factors. Certainly it discredited much faith in reliance upon market forces to tolerate a disaster of this magnitude. In this sense, the experiment in attempting to modify or negate market forces was induced by the depression. The form it took in the U.S., namely development of what came to be called the “fourth branch of government,” (see below) was relatively benign when compared with the more radical experiments with central planning, widespread nationalization and the like elsewhere.

But in the bastion of free enterprise, private property capitalism and democracy, it was not clear what should or could be done. The response in transportation was essentially more of the same. It was argued that the ICC failed earlier because it did not have enough authority and jurisdiction over all modes especially with the rapid growth of bus, truck and automobile traffic stimulated by the burgeoning highways. The quality of the Commissioners and staff was likewise not adequate. Since reliance on market forces seemed out of the question in the circumstances of the depression, it seemed logical to remedy these defects by legislative and personnel changes. After such measures, the ICC could reduce the reliance upon competition, regulate entry and exit, determine mergers, set rates and fares, monitor investments and financial arrangements of the carriers, specify accounting practices and

so on in an effective fashion. In short, it could coerce or induce the modes to behave like a more effective cartel that provided many social-type services, such as special rates for agricultural produce, low passenger fares for the poor, service even at a loss to outlying communities without alternative means of access, and so on. These services could be financed by cross subsidy from profit-maximizing rates and fares on other traffic. Overall profits could be limited to some reasonable amount to retain an aura of respectability and attract additional private capital when needed. The whole procedure would impart a degree of stability, permanence and, it was hoped, adequate level of service that would keep costs low and passengers and most shippers happy.

Transportation is tailor-made for this type of regulation. Each firm is a multi-product enterprise. Railroads produce not only passenger service and freight but within these categories the shipments of each commodity and travel of each passenger class between each pair of points or origins and destinations, can be viewed as a separate submarket having its own more or less specifically assignable costs and demand function. The game is to decide which services should be provided at below or close to avoidable cost, for whatever public purpose seems relevant at the time. All other prices would be set to achieve enough profit in the aggregate over and above these losses or reductions in net revenue to yield a fair, normal or otherwise "reasonable" rate of return on the total assets devoted to transport.

This is a system of discrimination designed to accomplish selective social and economic purposes at least cost. Private ownership and operation are believed to ensure the latter while economic regulation implements the social purposes responsive to Congress which creates the agency, determines its objectives and often its *modus vivendi* as well.

As noted, transportation is a natural since it has all but one of the ingredients necessary for effective price discrimination, namely, markets that are readily separable and demand elasticities that differ substantially in each submarket. For example, rate differences unrelated to the costs of transport can exist for different commodities because shippers cannot convert a high rated good (say, finished automobiles or diamonds) into a low rated good (say coal) to take advantage of the lower rate. Thus, freight classification systems early emerged in rail transport with rates applying to what was being shipped. It did not have to be this way. Shippers could have offered

packages of "things" to be transported and have rates based strictly on weight and distance without concern for the contents of the containers (i.e. the specific commodities involved). Indeed there are some so-called FAK (freight all kinds) rates and there is now much containerized shipping. But rate structures developed around classes of commodities often finely distinguished because this was far more profitable since it permitted more extensive discrimination. Thus, rates are based upon commodities (their values, shipping characteristics such as density, fragility, etc.) and specific origins and destinations, thereby permitting a vast array of different mark-ups over cost. For each such movement there will be an elasticity of transport demand that depends upon the product of the elasticity of demand for the commodity in any particular market and the ratio of the freight rate to the delivered value of the commodity. Since commodity demand elasticities vary sharply, not only among commodities and in particular markets but also over time, and freight ratios do likewise, it is clear that *transport* demand elasticities are also highly variable and differ among the various submarkets. This permits substantial price discrimination or large differences in the mark-ups over computed marginal costs.

There is more to the story but the main point is that for suppliers of transportation services the opportunity to discriminate and hence become more profitable is uniquely available compared with most other producers. In addition, it can be argued that railways have more incentive to discriminate because of their cost structure and behaviour but that is yet another story.⁴⁸

However, the foregoing suggests that only two of the three preconditions for successful price discrimination naturally exist in transportation. The third is the requirement of some degree of monopoly power. Without the ability to prevent entry of rivals in any market, it will not generally be possible to maintain a rate or price much above marginal or out-of-pocket cost without formal collusion. In this sense competition is said to be the great leveller. It forces all prices toward the marginal costs of the least-cost producer and precludes much discrimination. Although rails at one time had a large degree of monopoly power in numerous submarkets which permitted them to develop an elaborate classification system, this began to crumble and was extensively undercut by the development of bus and air passenger alternatives and especially trucks which could pick away at the upper end of the rate structure earlier put in place.

Maintenance of the discriminatory system in the face of growing shipper and traveller options, required that the common carrier segment of transportation be provided with monopoly power wherever possible through regulation or protected from freedom of entry into the high-yield submarkets. Thus regulation had to be tightened in the late 1920s and especially the 1930s to prevent the destruction of such discrimination. Regulatory policies restricting new entry into particular markets, specification of *minimum* rates and fares and other efforts designed to restrain competition were needed, not to protect the public as often believed, but rather to protect the carriers and especially the *system*. This would allow the implementation of a wide variety of social and other purposes unrelated to efficiency of production or allocative efficiency among modes or carriers. Regulation, if successful, would permit substantial mark-ups over cost in some submarkets that would provide higher net revenues. These in turn would allow much lower, and even below cost, rates and fares in other submarkets to effectuate some predetermined “social” or special purpose while maintaining private ownership and operation since overall revenues could, it was hoped, be made to exceed total costs by amounts yielding fair or normal rates of profit. Thus, cross subsidy in the absence of competition through regulation came to prevail in rail transportation and was even sought to be applied to bus, truck, air and inland waterway transportation.

The depression raised serious doubts about the efficacy of a policy based on competition and market forces. Indeed it is surprising that, with all the excess capacity resulting from the collapse of production and incomes, more cartel-like arrangements along the lines of the *National Industrial Recovery Act* codes of “fair” competition were not established. More surprising is that overtly socialist or communist measures involving widespread nationalization or more radical changes in the U.S. economy were not attempted. As it turned out, commission-type economic regulation was expanded within transportation and in various other areas as well. In fact the 1930s became the hey-day of the regulatory commission despite the dismal show put on by the ICC earlier. So strongly was there a perceived need to extend regulation that it was rationalized in terms of what was referred to as the “fourth” branch of government, alongside the executive, legislative and judicial. In a real sense, extended regulation was the U.S. alternative to socialism, communism and more detailed central economic planning at the microeconomic level. Keynesian economics later turned out to be the

alternative at the macroeconomic level. The former failed long before the latter although this was not recognized, nor were the linkages between the two appreciated until later in the 1970s.

So important are the 1930s in the evolution of United States transportation infrastructure and policy, that the changes that took place then and their cumulative effects became the focus of the attacks in favour of market forces of the 1960s. The remainder of this section will examine briefly the road to more effective cartelization and further retreat from efficiency criteria.

TRANSPORTATION INFRASTRUCTURE

The big change in infrastructure policy relates to the highways. The policy had been one in which limited federal aid was given to the states for pre-determined roads and in amounts strictly limited by federal budgets. The states and local communities were expected to finance the bulk of total construction costs and all of the maintenance costs. There were even questions regarding the constitutionality of any federal aid to highways, a viewpoint that dates back at least to the days of Presidents Monroe and Jackson. Yet the logic of increased federal participation was inexorable. Without federal aid there could have been no national system of interstate highways, such as now exists, although even such a system was questioned as late as the 1950s and failed to pass Congress until President Eisenhower referred to it as the national system of interstate and *defence* highways. Beyond defence, the federal government also has a duty to promote the free flow of interstate commerce which virtually cried out for increased participation in the slowly evolving road system.

All this timidity concerning federal aid changed with the Great Depression. Federal caution on highway spending was abandoned. The economy had to be stimulated no matter what. Since this required increased spending, it seemed prudent to focus on something useful rather than strictly make-work projects. Throughout the 1930s, therefore, highway expenditures were part of the strategy to promote economic recovery. Federal aid was extended well beyond the confines envisaged by the highway Acts of 1916 and 1921. By 1936 the Chief of the Bureau of Public Roads could report about the previous three years that "all highway construction was administered with employment of those on relief rolls as the primary objective."⁴⁹ This enormous change in policy involved a massive increase in federal

outlays for roads, which rose from barely \$94 million in 1930 to over \$1,170 million during 1939. Some 80 percent of the total in 1939 consisted of special federal relief program outlays. By the end of the 1930s state and local governments supplied fewer funds for construction than they had at the beginning. The former relationship among federal, state and local governments, so painstakingly developed since 1916, had been thoroughly transformed as had the motivation for highway expansion. As a result, the nation emerged from the Great Depression with a vastly enlarged and improved highway infrastructure.

With this infrastructure, a thriving truck and bus industry arose. Our concern is not directly with the trucking industry except to the extent that it sharply reduced the ability of the railroads to continue subsidizing intercity rail passenger traffic to the point where they ultimately required relief which much later took the form of Amtrak and public subsidy. Competition from intercity bus lines and of course vigorous growth, even during recession, of private automobile travel made further inroads on the railways.

The trucking industry had another influence upon bus service since regulation of trucking was virtually demanded by the railroads after they became aware of the magnitude of the threat. Other interest groups, including the larger trucking companies, also demanded trucking regulation, complaining about excessive competition from irresponsible individual truckers who drove unsafe vehicles, were uninsured and otherwise reprehensible. Indeed, the trucking industry became a kind of "haven for the unemployed" who entered the industry in huge numbers buying some kind of vehicle on vendor credit, as a stopgap until other job prospects emerged. Such marginal operators were willing to carry any and all freight at rates down to or below out-of-pocket costs plus a limited amount for their own subsistence. The poor service and financial weaknesses of such operators did nothing to improve the shippers' image of the industry and gave even stable companies a bad name, reducing earnings to boot. Support for regulation of the trucking industry grew. Even an industry that possessed most of the technical features of being highly competitive was grist for the regulatory mill ostensibly to permit closer coordination and reduce the wastes of competition and instability.

On the other hand, the bus industry had achieved a high degree of stability by the mid-1930s possibly because of relatively greater railway involvement in helping create the large Greyhound system and somewhat more careful

state regulation because of the safety issue than applied to independent trucking. Most states sought to regulate bus transport as a public utility and reduced competition and new entry on many routes. Regardless of the differences, the attempt to provide the ICC with expanded powers to implement the "grand regulatory design" noted earlier, swept up both truck and bus transportation within the same legislation, examined in the following section. Suffice it to note here that the new highway infrastructure along with the previous and continuing advances in vehicle motive power, design, safety, speed dependability and comfort, utterly transformed the competitive and institutional structure of transportation in the U.S. Despite this transformation, the regulatory regime imposed was based upon the 1887 *Act to Regulate Commerce* and failed to recognize the structural differences between railroads on the one hand and bus, truck and air transport on the other.

Airline infrastructure also gave rise to virtually a whole new industry which now dominates the commercial provision of intercity passenger traffic. Unlike with rail and highways, the provision, operation and maintenance of the airway system are a function of the federal government and have been since the beginning. In the 1920s the U.S. Army engaged in aerial map-making and airmail experiments in conjunction with the Post Office. Congressional appropriations financed a system of radios throughout the country and developed a night airway system coast to coast before the end of 1926. The *Air Commerce Act* placed responsibility in the Department of Commerce to foster air transportation, developing and establishing of airports, civil airways and other navigational facilities. Responsibility was again transferred under the *Civil Aeronautics Act of 1938* which began the direct economic regulation of the airlines as well. The airways remain a responsibility of the federal government and subject to annual appropriations plus, more recently, receipts from airway trust funds analogous to the highway trust fund established earlier.

Airports are also publicly owned but usually by local or regional "port authorities." Although prohibited from building airports, the federal government nonetheless made funds available during the Great Depression from various public works and Works Progress Administration programs. Thereafter the *Civil Aeronautics Act of 1938* created a plan of regular federal participation which accelerated after World War II. Suffice it to note here that the infrastructure of this future strong competitor with rail long-distance travel received substantial federal assistance from the beginning including,

in the early days, direct mail pay subsidies for airline operations and, of course, enormous support in developing airplanes and preparing them for the transition from military to civilian and commercial.

Airline passenger growth was very rapid in the 1920s and continued throughout the Great Depression. By 1930 the number of airline company passengers had risen to 383,000 from 48,000 in 1928. The number passed one million in 1938 and reached about four million in 1941 before the war curtailed civilian travel.⁵⁰ Rail passengers averaged over 450 million during the 1950s⁵¹ so quantitatively the impact was insignificant. Qualitatively, the picture was different and became far more so after World War II.

Finally it should be noted that while road and air transport infrastructure and its use had expanded enormously, that of rail shrank both in mileage, mostly of branch line abandonment, and certainly in use. The latter was only temporary as World War II was to change all that if only for what was to be a fleeting interlude.

TRANSPORTATION POLICY

With the tremendous increase in transportation capacity in all modes due to the severity of the decline in aggregate demand, and government investments in non-rail infrastructure, public policy concerned itself with preventing further price declines and in fact trying to engineer price increases to prevent profitability from disappearing altogether. This is understandable since macroeconomic policy was then unheard of. In fact this approach had at least one element of macro policy, namely, that if profitability could be restored, private investment might be stimulated which would expand output and also employment. The inconsistency here was not perceived. Why would firms, even if restored to profitability by price increases, expand capacity when there was already too much of it and how could the prices be maintained in the face of increased output? However, the prevailing philosophy was that competition was excessive and had to be carefully controlled. This required extensive expansion of regulatory powers in those industries already regulated, namely, transportation under the ICC and power under the Federal Power Commission, and extension of regulation to other industries as well as pushing the NIRA codes of "fair" competition everywhere else as part of President Roosevelt's New Deal program.

Even before the depression, people had been encouraged by the effects of business-government cooperation during World War I and especially the Railroad Administration's success at unravelling the congestion and improving coordination among the carriers. The era of "scientific" management had already arrived. In addition, Thorstein Veblen's emphasis upon giving power to engineers to run the industrial plant in the interests of physical coordination, production of "serviceable" products and efficiency, a form of technocracy, rather than businessmen seeking profits, created something of a stir, as did Veblen himself.⁵² The point is that the reaction against markets not only took the form of reemphasis upon regulation and its extension but on the importance of "experts" on the regulatory commissions who had intimate knowledge of the industry just as engineers would know more about production than businessmen. Furthermore, since the regulatory goals of the firms would not be maximum money profits, production decisions would be less concerned with mere pecuniary values than with producing real, tangible commodities that were serviceable. The expert-dominated commission became the ideal. Commission-type authorities spread beyond transportation and power to other areas and included the Securities and Exchange Commission, the Federal Communications Commission and the National Labor Relations Board.

As noted, these sentiments predated the Great Depression but were vastly enhanced by it in attempts to protect business from unprecedented overcapacity. "The same premises that underlay the Keynesian revolution in macroeconomics — that government should take a hand, that expert public servants were more likely than business executives to choose the wise course — also applied to microeconomic regulation."⁵³

The arguments in favour of regulation and especially industry-specific regulation by a small group of people with special knowledge or expertise about the industry were not only economic — that is, justified by market failure, economies of scale, the so-called "wastes" of competition, natural monopoly, large-cost indivisibilities necessitating discriminatory pricing, the need to promote the industry and the like — but also political.⁵⁴ The extension of regulatory authority was rationalized in terms of need, inevitability and legitimacy — the "fourth" branch of government. While the role of government in an industrialized economy must increase, the existing branches of government — legislative, executive and judicial — were not designed to cope with it and its drive toward monopoly power and general cyclical instability. The rise of a specialized administrative branch was therefore

a logical next step in a political democracy. The general argument is not unlike that used by Japan to justify and rationalize the activities of the Ministry of International Trade and Industry (MITI). Indeed the kind of “guided capitalism” in the form of industrial policy that MITI exemplified in its early decades closely parallels that of the “grand design.”

REGULATORY EXPANSION

The ICC, after acquiring additional authority from the *Transportation Act of 1920*, sought to extend it even further as the motor carrier threat loomed larger. It was not however until 1935 that the *Motor Carrier Act* was passed which extended the Commission’s authority to regulate the trucking industry. The Act became part II of the *Interstate Commerce Act*. It had been strongly supported by the Commission, the Federal Coordinator of Transportation, the railroads and the larger organized trucking firms, but opposition from motor vehicle manufacturers, shippers and small truckers fearing that regulation would restrict the industry in favour of railroads held it up for many years even after the depression was well under way. Most economists opposed it on grounds that the motor carrier industry was about as far from a natural monopoly as could be conceived. Entry was easy, fixed costs were low, vehicles could readily be shifted to accommodate traffic flows and adjusted to total business, no scale economies were evident, and so on. But the economists’ hour had not yet come. The grand design needed to be tried. So the Commission took under its wing an industry almost the complete opposite of the railroads.

The ICC’s workload burgeoned. By November 1, 1936, there were some 86,000 applications for certificates and permits, one half of which were protested; over 53,000 tariff publications; almost 17,000 schedules and 2,000 contracts filed. The Commission was swamped. In addition the Act required fine distinctions to be made among common, contract, exempt and private carriage, what constituted an “agricultural product” since these were exempted from regulation, and so on. More important, the Commission had to determine what the “inherent advantages” were of truck vis-à-vis rail and how to “preserve” them.

As if this were not enough, the *Transportation Act of 1940* gave an already swamped and confused Commission jurisdiction over coastwise, intercoastal, inland and Great Lakes common and contract water carriers in interstate and foreign commerce! All of this and more was bestowed under the statement

of national transportation policy noted in Section I and also noted for its ambiguity, inconsistency and requirement to meet the “needs of commerce, of the Postal Service and of the national defense.”

There was some discussion about placing air transport under the ICC’s jurisdiction as well but a certain amount of common sense prevailed and consistent with the regulatory mania, a new commission, the Civil Aeronautics Authority, (later Civil Aeronautics Board) was created in 1938.

By 1940 therefore the ICC had been granted abundant authority to rationalize or coordinate the several modes of transportation, except for air, and including both freight and passenger transportation without distinguishing between them. The Commission had been powerless to do anything in the face of the depression. It certainly could not guarantee rail profitability, the rule of rate-making notwithstanding, and was unable to obtain much in the way of inter-railroad coordination or cooperation either even under those critical circumstances. It was now enjoined to seek rail-truck-bus-water, etc. integration. During the 1920s, the ICC had performed dismally. Having failed to lead for over 20 years, the future did not look propitious no matter what authority the agency possessed. The “grand design” for an important segment of the newly rationalized “fourth” branch of government did not appear to be in very good hands unless some more enlightened and energetic commissioners could be found.

The outlook for the newly formed CAA was more propitious if only because it seemed to start with a blank slate and fresh faces. However, the principles expressed in the new Act were like those of the *Act to Regulate Commerce* of 1887 and the recently passed *Motor Carrier Act*, namely “the usual system of economic regulation” as the Senate Committee on Commerce explained. This meant that the industry was to be treated as a public utility currently subject to excessive instability, cut-throat competition, inadequate earnings and the usual excuses for bringing regulation to bear. In this case of course, as with rail, trucking and the bus industries, the major problem was the Great Depression. Few industries or even firms outside of transportation fared very well nor could be expected to do so in the presence of enormous excess capacity. What is not so evident, is the degree of undesirable instability that might have occurred in more prosperous and less unusual and turbulent times. However, when recovery came about there was ample opportunity to test the feasibility of the regulatory regimes put in place in transportation. This was to come after World War II.

4. WORLD WAR II TO THE *TRANSPORTATION ACT OF 1958*

Massive increases in federal government expenditures, mostly in preparation for the war then raging in Western Europe, occurred in 1939 and accelerated rapidly following U.S. entry into the war after December 7, 1941. This was the stimulus the economy needed. Real GNP surged by almost 8% per year during 1939 and 1940 and by 18% per year through 1944 — incredible growth rates by contemporary or indeed any standards. Such growth of course was possible because of the enormous excess capacity during the depression. The unemployment rate fell from a whopping 25% during 1933 to less than 5% in 1942 and below 2% for the next three years. It was as if the country were a coiled spring waiting for some catalyst to let it loose. The war did so with a vengeance.

Such dynamism had enormous impacts on almost all aspects of the economy and society but few more dramatic and positive than on railroads. On the verge of almost total financial collapse in 1938, rail tonnage and revenue records were broken during 1942, 1943 and 1944. Passenger revenues hit an all-time high in 1944 and the passenger deficits turned into substantial surpluses during 1943, 1944 and 1945. Railroads carried “97 percent of troops and 90 percent of army and navy supplies moved within the United States.”⁵⁵ The railroads were indeed riding high. Part of this was, however, due to curtailment of competition. Because of fuel and equipment shortages, truck and private car travel was sharply curtailed. Bus traffic increased somewhat but civilian air travel was virtually eliminated. No wonder the public, industry and the military flocked to the railroads who literally never had it so good.

The euphoria for the railroads was short lived. The end of the war meant an end to massive troop and armaments movements. The dismal passenger service was well remembered by those subject to it during the war and when restrictions on air and automotive travel were removed and production of new and improved planes and cars began, people and freight left the rails in droves. The long steady relative freight decline resumed with the rail share of total ton-miles falling from 62 percent in 1939 to 49 percent by 1956 and continuing thereafter.

Even more dramatic was the resumption of the *absolute* decline in passenger traffic. Almost monotonically, passenger-miles fell from 95 billion at the wartime peak in 1944 to barely 23 billion in 1958, with a further steady

decline to less than 11 billion in 1970, the year before Amtrak. The number of passengers fell almost as dramatically. More significantly, rail passenger traffic became minuscule in proportion to total intercity passenger-miles, including private automobiles, falling below four percent and a diminishing share of common carriage already down to less than a third by 1958. The war years were but a brief upward blip, a mere interlude, in the long-range decline of both rail freight and especially passenger traffic.

More significant for the railways were the burgeoning passenger deficits which reached alarming highs of over \$700 million by 1957. Despite more or less valiant efforts of the railways to employ more sophisticated and better equipment and upgrade service quality, they were no match for the service quality-cost alternatives provided by automobiles and airplanes, assisted to be sure by a public sector prone to concentrate large amounts of resources steadily into improvements in their respective rights-of-way. Intercity rail passenger service seemed doomed to chronic unprofitability.

The railroads did give the passenger business a decent try in the post-war period, much better and more sustained than the “streamliner” gimmick of 1934. But the passenger business is different from freight, and since it was producing negative net receipts while freight traffic was at least somewhat profitable, and since passenger traffic had traditionally been viewed as a by-product, often expensive but occasionally exotic in some aspects of its service, there was little more that could have been expected from rail management. There is no evidence of deliberate downgrading of service to induce customers to stay away because until 1958 there was no certainty that states, which had sole authority over discontinuance, would permit abandonment. In fact there was much evidence to the contrary.

However there came a point after which, despite some real efforts by management to change the cost-quality options, the profit prospects seemed hopeless. The Hosmer Report, examined below, also concluded that the passenger decline could not be attributed to rail management hostility or indifference, arguing that in fact they had made a “noble” effort to preserve the service out of pride and public obligation, but that even a marginally profitable outcome was hopeless.⁵⁶ At that stage and having dissipated much of the huge wartime surpluses in attempts to improve service, the railroads threw in the sponge and simply let service and equipment deteriorate. This phase provides most of the anecdotal evidence of deliberately

poor service, rudeness of personnel, indifference and so on.⁵⁷ Hilton is probably correct when he notes that “by 1958 railroad passenger service had demonstrated itself to be the most uneconomic activity ever carried on by private firms for [such] a prolonged period.”⁵⁸

ICC POLICY

From 1940 when the Commission was given substantial authority over all modes of transportation except aviation, and presumably had developed some concept along the lines of the “grand design” noted in Section III, nothing positive was done until after the war. During the war, President Roosevelt by-passed the ICC and created a separate Office of Defense Transportation (ODT), much as President Wilson had ignored the Commission in the earlier war, though not resorting to federal control. Apparently the Commission was not held in high enough regard to coordinate the transportation industry during either war even though it had jurisdiction over by far the largest proportion of transportation assets, and in the *Transportation Act of 1940* was charged with coordinating the transportation system. To be sure, Joseph B. Eastman was chosen to head the ODT but not because he was an ICC commissioner but because he was a man of exceptional talent and leadership abilities. His death in 1944 was a serious loss to the Commission since no commissioner “of comparable talents rose to take his place.”⁵⁹

It was not therefore until the post-war period that the Commission had the opportunity to see what it could do about improving the performance of the transportation industries through attempts to implement its grand design. The results speak for themselves. By the mid 1950s the railroads faced financial disaster again. Part of this was due to ICC policies that have regularly been viewed as inconsistent, vacillating, uneconomic and otherwise misplaced or foolish. I submit however that the Commission was attempting to implement a larger vision of the mandate of the *Transportation Act of 1940* than perhaps Congress had in mind.

Certainly the Act was interpreted by the Commission as involving a multiple set of objectives that made enforcement difficult. For example, the Commission asserted that in any important case it must consider the “economic effects on all shippers, on towns, cities, ports and regions, on the carriers themselves, and, of course, on the consumer”⁶⁰ in addition to the “needs of the commerce of the United States, of the Postal Service,

and, of the national defense" as the *Transportation Act of 1940* put it in its policy statement. Without having some notion of the Commission's relative weighing of these several objectives, it is not possible to judge its decision in any case. Worse still, since the Commissioners themselves have never provided any information on the importance they attach to any specific objective, it is clear that every decision is "justified" or justifiable under some set of weights that seemed to shift from case to case. Furthermore, there is no presumption that the objectives are economic in nature despite the adjective "economic" in the foregoing. Thus Commissioner Webb, in referring to the frequent use by the Commission of average costs containing many items that are arbitrarily averaged into any given figure and hence having no relevance whatsoever to what economists mean by "marginal" cost, explained that:

the allocation of constant costs to specific traffic . . . is essentially arbitrary . . . There is no doubt in my mind that such fully distributed cost constructions are bottomed on economic fallacy. . . [But] this form of economic nonsense . . . may be entirely sound from a regulatory point of view . . . [because] regulation is designed to achieve a number of important objectives, the value of which cannot be determined on purely economic grounds.⁶¹

This non-economic aspect goes back to the beginning of the Commission. In its first annual report, the ICC stated that "[i]t was . . . seen not to be *unjust* to apportion the whole cost of service among all the articles transported, upon a basis that should consider the relative value of the service more than the relative cost [emphasis added]."⁶²

There are complaints that the ICC "has been on all sides of the ratecutting question" because in one case it required rail rates to be raised "to match those of their river-truck competitors" while in another it asserted that "no carrier should be required to maintain rates which would be unreasonable, judged by other standards, for the purpose of protecting the traffic of a competitor," and in another case that "a rate should 'not be lower than necessary' to secure a fair share to traffic." These decisions do not mean inconsistency,⁶³ rather they reflect ICC's multiple objectives. Higher rates above rail costs may be justified on the grounds of protecting the traffic of other common carriers to keep them in business in the event of defence needs and to preserve their "inherent advantages." Rates closer to costs

but still compensatory may be justified in competitive situations if they retain or attract traffic from less regulated or exempt carriers.

Any rate change is bound to have differential effects on ports, other carriers, consumers, national defence and so on depending upon the situation of the specific case. Thus it is unfair to castigate the Commission for its case-by-case approach and failure of enunciate a uniform policy applicable to all cases. To be sure, the Commission's approach creates a lot of confusion and uncertainty as to the outcomes of any case. This also causes much delay, as well as income for lawyers and consultants of various kinds as each case is explored in all of its dimensions by all interested parties. Most of the delay is a product of the legalistic approach required by the *Administrative Procedure Act* (1946) but it is also implicit in the ICC's view that each case is virtually *sui generis*.

The Commission supported legalization of railroad rate bureaus which finally became law over President Truman's veto in 1948. In many ways the ICC sought to have the common carriers, regardless of mode, behave like a multi-enterprise cartel dedicated to the public interest though still under private ownership and management but influenced by regulation administered and interpreted by the ICC. In specific situations involving rail and truck competition, the Commission would apply a "share-the-traffic" criterion if the bulk of the rail and truck traffic involved common regulated carriers and would prevent rate competition that would jeopardize the net revenues of either mode. On the other hand, if there was actual or potential non-regulated (private, exempt or even contract) carriage, the Commission would permit the more regulated carriers to slash rates to levels barely above out-of-pocket costs if necessary, in an attempt to maintain at least some profitable *common* carriage in the particular business.

This was all part of a vain attempt to prevent the steady, increasingly evident erosion of regulated common carrier traffic to private, own-account, exempt or less regulated business. Other attempts involved increasingly rigid definitions over what constituted permissible non-common carrier status, such as reducing the number of contracts allowed to a contract carrier before its status became de facto "common" and hence subject to the full panoply of ICC rules and regulations; narrow definition of exempt commodities under the agricultural or bulk commodity exemptions in the laws applicable to truck and water transportation; and the extreme

“poisoned vehicle doctrine” that held that any specific *vehicle* that had ever carried an exempt commodity was forbidden from ever again carrying non-exempt freight!

The totality of Commission activities between 1940 and the mid 1950s suggests that interpreting the vision as a “Grand Benthamite Design”⁶⁴ or simply “grand design” as mentioned earlier sheds some light on what the Commission did or was trying to do. The argument goes as follows:

Rate regulation as practiced by the ICC, if completely effective, would maintain higher rates on specific traffic than would otherwise occur, would involve greater rigidity, tend to preserve a demand-oriented rate structure, reinforce misapplication of value of service pricing but emphasize commodity discrimination as the most effective way of covering the nonassignable costs, frequently allow more than one mode to participate in a given piece of business regardless of relative costs, consider a wide variety of noneconomic objectives and externalities, treat each case as *sui generis*, more effectively prevent rates from falling below marginal cost, and would force a kind of public accountability and greater cost consciousness upon the common carriers. Insofar as one can impute rationality to contemporary regulation, the overall impression is that the Commission seeks a rate structure such that all nontraceable costs including those additional costs arising from pursuit of noneconomic objectives and the added duties of common carriage are recouped from shippers on the basis of willingness and ability to pay; that is to say, the markup over computed out-of-pocket costs would be determined by the relative level and elasticity of the demand for transport of particular commodities. Individual rates would be adjusted from the maximum profit level by consideration of external economies and diseconomies as far as regulated carriers are concerned and overall profit limitation is implied. Rate regulation, if completely effective, thus attempts, albeit in a crude and implicit way, to assess the benefits to the regulated carriers as a group and weigh these against whatever externalities appear to emerge in particular instances. It is true, of course, that, unlike Bentham, the ICC does not explicitly add up the net benefits to each shipper, carrier, and community involved or affected and then strike the balance, approving those rate changes for which the sum of the net benefits exceeds zero and disapproving all others. But the frequent

references to injury or noninjury to some shippers, regions, defense interests, or other carriers due to proposed rate changes, suggests that an interpretation of rate policy in terms of Bentham's felicific calculus may not be so farfetched as it seems on the surface.⁶⁵

Granting for the moment that this was the (mostly implicit) vision that derives from ICC actions, it still does not exempt the Commission from charges of failure to promote coordination among the modes. Nor is it exempt from charges that its activities, however well-intentioned, did little to preserve common carriage. In fact, it helped reduce common carriage which in many respects had become somewhat obsolete anyway. It is also true that the railway situation would probably not have been significantly different regardless of the Commission's actions. The forces of technological change that so improved the operations of the railroads' principal rivals and the massive support given by the federal government for right-of-way and technical development of the vehicles, especially aircraft and its close association with the military, would have led to the relative demise of rail freight traffic anyway and the absolute demise of passenger traffic too, in fact much sooner, had efficiency criteria prevailed or been influential earlier. Nor can one avoid the conclusion that the "grand design" was enormously wasteful in economic terms, of which more later.

THE WEEKS REPORT

In any event by the mid-1950s another rail crisis loomed. Since part of the railroads' problem was associated with regulatory policies of the ICC about which something might be done, President Eisenhower appointed a Presidential Advisory Committee on Transport Policy and Organization with Sinclair Weeks, the Secretary of Commerce, as chairman. It was popularly referred to as the Weeks Report.⁶⁶

The Report, released on April 18, 1955 was primarily about railroads and freight. However three of its main items have relevance to passenger travel by rail and other means and indeed were influential in the *Transportation Act of 1958*.

Two views prevailed at the time concerning what to do about transportation policy in general:

- (1) extend the authority of the ICC because much of the malaise was the result of incomplete jurisdiction. In passenger travel, the Commission had no authority over aviation, private transportation by auto and intra-state bus travel. In freight, regulation was complete over rail but barely one third of truck ton-miles and between 10 and 15 percent of inland water transport were subject to ICC regulation. For more complete coordination, it was argued, exemptions from regulation needed to be reduced. This applied mostly to private and exempt trucking where the Commission regularly tried to construe the extent of exemption narrowly. This approach was obviously mostly favoured by the Commission itself and those who believed that more regulation was essential to sustain common carriage, a transportation system constantly prepared for defence and a stable, dependable and ubiquitous system regardless of concerns for costs or profitability. In short, transport was viewed as a kind of public utility, available at most or all times between most or all places. It was a kind of entitlement and the "right to mobility" something akin to one of the basic freedoms.
- (2) the other approach was to rely more on market forces and progressively strip the regulatory commissions of their authority. This was based upon two main premises: one, that the transport revolution over the past 50-plus years had rendered transport markets for both people and goods sufficiently competitive that public control, beyond the antitrust laws, was unnecessary and, two, that even if this were not totally the case, the existing system of regulation was so cumbersome, costly, unwieldy and unfair that it ought to be scrapped anyway.

The Weeks Report had elements of both of these views. Recognizing that the 1940 statement of national policy had been interpreted by the Commission as a directive to allocate traffic among carriers on its own discretion rather than via market forces, the Report stresses that a veritable transport revolution had taken place. "The net result is a competitive system . . . that has eliminated the monopoly element of . . . some thirty years ago." However, government has kept "and in fact, intensified its regulation." The Report therefore came down strongly in favour of "increased reliance on competitive forces . . . in rate-making," one of the first of such utterances to emanate from a prestigious government source.

However, the Report could not break away from the past completely and recommended that rate competition take place only within a range

determined by the ICC. This kind of ambivalence persisted throughout; namely recommendations that would permit a more market-oriented approach on the one hand while, on the other hand, strengthening or broadening the scope of regulatory authority. For example, to help maintain, however artificially, a common carrier industry, recommendations were made to narrow the extent of exempt and private carriage in trucking. Defence issues also loomed large with great stress placed upon maintenance of excess capacity which involves an inconsistency since reliance upon competition would preclude persistent excess capacity. It is impossible to maintain a strong competitive common carrier industry and excess supply in the event of hostilities unless the excess is kept out of commercial competition which is not the tenor of the Report.⁶⁷

The main aspects of the Report and the resultant *Transportation Act of 1958* of relevance to passenger transport include those items placing more reliance on market forces and those moving further toward regulatory control. The latter would tend to preserve rail intercity passenger service and the former reduce it. The Report stressed that "losses incurred from passenger service . . . must be borne from earnings realized from freight service. . . . Thus railroad shippers . . . are being required to subsidize . . . those who benefit from the utilization of passenger train services." This was a strong plea to place rail passenger abandonment in the hands of the ICC and away from the states and make the Commission decide when such deficits impose an "undue burden upon interstate commerce." This was a tepid move toward market forces since it did not permit discontinuance upon the showing of unprofitability alone but dragged in concerns of "undue burden." Ironically, the original draft of the *Transportation Act* had provided that any demonstrably unprofitable passenger service might be discontinued. The removal of this provision for reasons relating to fear of elimination of suburban services in New York City, left the ICC without any specific criterion and, thus, free to decide such issues in conformity with its "grand design." On the other hand, recommendations for more rate freedom, even if only within a range specified by the Commission, would stimulate profits from freight shipments which would then permit continued subsidies to passenger service. Recommendations for federal loan guarantees for portions of rail plant and equipment purchases were also proposed and, other things being equal, this non-market effort would help support perpetuation of deficit passenger service.

Thus the Report was a mixed bag *en route* to liberalizing regulation. The resultant regulation in 1958 was equally mixed. Hilton, after an exhaustive study of it from the perspective of 10 years later concludes:

two provisions . . . move in the direction of market allocation of resources. Section 15a(3) [prohibiting umbrella rate making] is a halting effort at decartelization, and to the extent that Section 13a accelerated withdrawal of passenger trains, it furthered the working of market forces. Unfortunately, such benefits as society achieved from these provisions were to some degree counteracted by the Act's two provisions which tended in the opposite direction: the guaranty of loans worked contrary to market forces for disinvestment in a declining industry; and the motor carrier provisions were . . . a straightforward extension of cartelization.⁶⁸

THE HOSMER REPORT

Howard Hosmer was an ICC examiner who was directed in 1956 to look at the rail passenger deficit and suggest ways to eliminate it or at least reduce it to tolerable levels, presumably in the hope that present arrangements might continue or that feasible alternatives to preservation might be found. Clearly the crisis stage had been reached. As the ICC Report one year after Hosmer's put it:

The financial loss [from deficits] is real; it is large and appears to be growing; and it endangers the present and future welfare of the railroad industry. . . . [I]f this threat to realization of the objectives of the national transportation policy is to be substantially lessened, responsible efforts in this direction must reckon with facts as they are. If the statistical trends of 25-odd years prove any one thing, it is the folly of awaiting more fortuitous events. Though many people still prefer to travel by rail for one reason or another, they clearly constitute a minority — a minority that grows smaller with each passing year. Figures for 1958 show that for the first time in the 70 years of recorded history of the American railroads, their passengers numbered less than 400 million. This decline has occurred against a background of an expanding population with more time, means and the desire for travel. In terms of intercity passenger-miles, the 'travel market'

increased almost 60 percent from 1940 through 1957. Though the total 'for-hire' share rose by slightly less than 9 percent, railroad passenger-miles declined 27 percent.⁶⁹

There was certainly little time to lose. Though the ICC Report was in large part based upon Hosmer's, nevertheless it refused to accept his conclusions. Wishful thinking triumphed again when the Commission concluded that "We are of the view that the complete elimination of passenger-train service would not be a solution in the public interest. Economic railroad passenger service is . . . essential for the Nation's well-being and defense."⁷⁰ The Commission then enunciated 10 measures to somehow "save," and/or render economical, rail passenger service — only one of which was ever implemented: the 10 percent federal excise tax on passenger fares was repealed — not exactly a radical move. Nor did anyone suggest that even if all the measures had been implemented "economical" service was possible. Hosmer certainly thought not. We turn now to his analysis which is of relevance today.

After looking at the size of the deficit, he disposed of arguments that it was arbitrarily overstated by noting other studies arguing that the reverse was true. He deftly accepted the Commission's earlier finding that the figures, though based upon somewhat arbitrary but reasonable rules for separating expenses between freight and passenger operations "were adequate for the purpose for which they are intended," and turned to their causes.

On the demand side, he noted that the public at large had exhibited a distinct preference for air and automobile travel. Since the typical age of a rail passenger in 1958 was over 45, the bulk of the shift away from rail entailed more youthful riders whose future habits were unlikely to include rail after having become accustomed to the superior quality of the alternatives and who had by then gotten over any fear of flying associated with earlier air travel. Data at the time also suggested that the future of rail demand was severely limited because of several elasticity estimates which indicated that income elasticity for air and automobile travel was plus 2.5 and 1.2 respectively whereas that for rail was a *negative* 0.6. With rising incomes especially true in the United States of the 1950s and 1960s, the demand for rail intercity passenger service could expect to drop absolutely while increasing rapidly for its rivals. The rise in incomes also increased the value of time for travellers, especially business travellers, which generated even

greater relative advantage to air travel and frequently to automobile and bus as well. As Hilton put it, "Only low-income people, persons with an irrational fear of flying, or aged persons whose alternative uses of time are negligible are likely to opt for rail."⁷¹

On the cost side, Hosmer argued that rail transport was more labour intensive than its competitors and less subject to productivity improvement. These facts, combined with the stringent operating rules applicable to rail labour, would make their cost disadvantage even greater, thereby further accelerating their overall demise. As if to bear out Hosmer's analysis, a study in 1966 indicated that rail costs per passenger for a 45 percent load factor between San Francisco and Los Angeles were between \$18 and \$25 whereas the air fare was less than \$10 on one of the airlines and slightly over \$9 by Greyhound Lines.⁷² Since the rail fare had to lie somewhere between the other two because of perceived relative travel preferences, rail losses were inevitable, especially at the actual fare of \$12.50.

Thus, given the apparently inexorable or, at least, most likely trends, the rail intercity passenger business was economically hopeless. Hosmer stated this without equivocation:

For more than a century the railroad passenger coach has occupied an interesting and useful place in American life, but at the present time the inescapable fact — and certainly to many people an unpleasant one — seems to be that in a decade or so this time-honored vehicle may take its place in the transportation museum along with the stagecoach, the sidewheeler, and the steam locomotive. It is repetitious to add that this outcome will be due to the fact that the American public is doing about 90 percent of its traveling by private automobile and prefers to do so. . . . If the railroad passenger-miles (other than commutation) continue to decline at the average rate of reduction between 1947 and 1957, the parlor and sleeping car service will have disappeared by 1965 and the coach service by 1970. . . .⁷³

The latter projections were almost precisely on track if one dates a new era for intercity rail passenger service in the U.S. from the formation of Amtrak in 1971 and the demise of most luxury services five years earlier. At least for the existing railroads, passenger service died as predicted.

Thus, on the eve of obtaining authority over passenger-train discontinuance, the Commission had at hand an analysis that would have provided an appropriate guideline for policy. This it chose to reject. While in the years following the *Transportation Act of 1958* many passenger trains were allowed to be dropped, the procedure was more prolonged, painful and costly than would have been the case had Hosmer's Report received the attention it deserved. The hopes of implementing the "grand design" more effectively despite such evidence presented by Hosmer of the power of market forces to effectuate change, made the Commission continue, like King Canute, to order the waves rolled back. Thus the railroads, despite growing rates of abandonment begrudgingly granted by the ICC, continued to suffer passenger losses ranging from a low of \$375 million to a high of \$524 million and averaging \$426 million per year from 1959 through 1970.

THE SITUATION IN AIR AND HIGHWAY TRANSPORT

While the railroads were having their problems and suffering from low earnings overall and continued miserable passenger performance, air and private travel reached new heights while bus service limped along. Economic regulation of air seemed at the very least to have accommodated such growth and perhaps helped it along marginally. Certainly public policy with respect to infrastructure mightily assisted all three. These issues will be briefly examined here in reference to the period before the heavy artillery was brought out to begin the attack upon the regulatory commissions in transportation following the timid beginning with the *Transportation Act of 1958* and the Weeks Report which were not even discernable salvos.

Air Transport

The growth of the commercial side of air transport was phenomenal. Employment in the industry increased from a mere 13,000 in 1938, the year of regulation, and by 1958 exceeded 150,000. Revenue passenger-miles leaped from barely half a million to over 31 billion between these years and carrier revenues from \$58 million to over \$2.2 billion.⁷⁴ Rate of return on investment for the carriers averaged over 10 percent per year during the 1950s.

The airway development proceeded apace without charge to users and with direct subsidies linked to mail pay administered by the CAB from 1953 to deregulation in 1978 and directed mostly to carriers serving small

communities. Airport development has accommodated and partly stimulated air traffic and carrier profits since state authorities and local governments generally build, maintain and operate the facilities at fees that normally do not cover operating costs and certainly not capital costs. In the new environment these implicit and explicit subsidies are being phased out.

Carriers acquired vast amounts of new and higher technology equipment. Ironically, in the year of the Hosmer Report and the *Transportation Act*, both of which involved the future of passenger travel by rail, jet aircraft were first introduced on domestic airlines thereby making Hosmer, at least, appear even more prescient. Newer, faster, more fuel-efficient and sometimes more comfortable aircraft have been introduced with singular rapidity ever since. The industry has been technologically extremely dynamic, perhaps too much so. In part the new technology improves quality of flight and this may have been excessively stimulated because the CAB early on adopted the policy of virtually eliminating rate competition among certificated carriers, forcing competition to focus upon the quality dimension. However, replacing an aircraft fleet with another more expensive before the first has been fully financed or paid for has often caused financial problems for the airlines along with fluctuating net earnings.

Suffice it to say that up until roughly 1960, U.S. airlines were dynamic, profitable and comfortable with both public *largesse* and direct economic regulation. In fact under the latter, no new trunk-line carrier had been allowed to enter the industry since the beginning of the CAB. Of the original 16 carriers in 1938, 10 continued to dominate the industry in the early 1970s, the other six having merged with the remainder. The CAB clearly "out-regulated" the ICC and in an industry where the workability of competition vastly exceeded that of railroads. Perhaps the earlier demise of the CAB in its entirety is just desert for cartelizing a workably competitive industry presumably in the public interest and after large amounts of subsidy seeking to promote it in the first place. The public paid at both ends.

Highway Transport

Bus transportation had not shown much dynamism through the 1950s. In fact, total intercity bus traffic (class I, II and III carriers) decreased slightly during the decade in terms of passenger-miles and fell behind air for the first time in the mid 1950s and, of course, has remained behind since. It

continued to dominate in number of passengers however, indicating the very short average length of trip, some 122 miles in 1983 and below 100 miles or less during the 1950s. The clientele come mainly from low income, non-professional groups, the relatively aged and very young, and whose travel purpose is primarily visitation of friends or relatives and sightseeing. These features seem to have characterized bus travel from the beginning and certainly since World War II, which probably accounts for part of the lack of dynamic growth of the number of passengers and passenger-miles over time.⁷⁵

The chronic high level of monopoly power in the industry may also have contributed to its sluggish level of output and relative decline. Although the operating units are such that a competitive structure would be expected to emerge, both early state regulation and federal regulation as part of the *Motor Carrier Act of 1935* succeeded in creating a high degree of concentration especially in a single company, Greyhound, which dominated the industry for over 60 years. Along with Trailways, the two companies accounted for over half of all intercity carrier bus-miles and over 60 percent of total industry revenue in 1976 and similar proportions in the earlier periods. It is an understatement to assert that "concentration is overwhelming in the bus industry."⁷⁶

Obviously for some submarkets or city pairs there may be effective rivalry from other modes, but in general it appears that regulation has helped to fashion another cartel out of a potentially workably competitive industry although there is no evidence that, following deregulation, the competitive nature of the industry has changed. The industry has simply not exhibited any signs of innovative development or dynamic growth even though its clientele has grown and presumably the number of tourist trips as well. There has been limited technical change in the basic vehicle and a pretty steady relative shrinkage in share of total intercity passenger-miles to insignificance. Regulation has also led to some cross subsidy with local service deficits being made up from small surpluses in intercity service including express and other incidentals.⁷⁷

Unlike rail passenger traffic, the bus industry does not appear to be destined for extinction. Although rates of return have decreased in recent years, they were not completely negative during the 1950s. Nor has the trend been downward for other indicators except in a relative sense. For thousands of

small communities, buses provide the only commercial access and egress and are of the sort that may well be suited to the non-business, non-professional patrons' needs, especially the elderly.

In the meantime private highway automobile travel continued to burgeon after the war. Fuel rationing ended, tires became available once more, Detroit geared up its formidable production apparatus and highway expansion and improvement went ahead with a vengeance culminating in the beginning of the Interstate System under President Eisenhower. Intercity auto travel rose from 383 billion passenger-miles in 1949 to 706 billion in 1960, and from 85 percent of the total to more than 90 percent. Automobiles improved markedly in styling, comfort, convenience and, to some degree, in safety and fuel economy during the 1950s although there was much more emphasis upon these later.

But the big events were the highways themselves. The National System of Interstate and Defense Highways, the "Interstate System," was created by the 1944 *Federal Highway Act* but funding for it was not approved until 1956. Two more highway Acts in 1956 set the system at 41,000 miles, later extended to 42,500 miles, and authorized new methods of financing both the Interstate and other federally aided systems. The Interstate is one of the largest public works endeavours in history. Ninety percent of the system is funded by the federal government and while its length only constitutes about one percent of total U.S. highway mileage, it accommodates over 20 percent of all highway traffic and connects virtually all cities (with populations of 50,000 or more in 1956) with limited access, multi-lane roads. The federal role also covers the ABC Program which includes primary and secondary highways and extensions into urban areas. The federal share of construction costs for this more extensive coverage began at 50 percent in 1956 but was raised to 75 percent in 1974, still somewhat below the 90 percent for the Interstate System. Overall federal expenditures for highway construction cover about 20 percent of the nation's total street, road and highway mileage.

A most significant aspect of the *Highway Revenue Act of 1956* was its creation of a Highway Trust Fund into which federal taxes associated with highway use are placed and the total each year allocated for the construction of federal aid highways. Before the passage of this law, highway expenditures had been paid from general revenues which included receipts from a wide

variety of taxes, levies, fines, fees and other sources of income to the federal government. The share that went to highways was then bargained for in a political process referred to as the annual budget. The trust arrangement, under which some or all of the proceeds of excise taxes on motor fuel, tires, tubes, as well as sales of cars, trucks, buses, trailers and heavy vehicles accrue to the fund, and total expenditures for construction are limited to such amounts per year. This is believed to have the advantage of assuring funds for future use since such tax receipts can be forecast with more accuracy than political bargaining over budget allocations on an annual basis. More important for efficiency purposes is that it provides a linkage between highway use and highway outlays. Since economic efficiency requires that prices charged for the use of publicly or privately provided rights of way should reflect their respective marginal social costs, this linkage is at least a step in the right direction.

For example, the favourable economic outcomes resulting from a largely procompetitive transport policy, require that suppliers of transportation services, as all other producers of goods and services, pay the full costs borne by society in producing such outputs. Even if the production of the output is for personal use, such as own-account automobile traffic, if the full costs of such production are not paid by the producer or immediate beneficiaries, economic waste will be the result. The producer, or in this case, the automobile user will have no incentive to "economize" on the number of trips made if each trip is not charged an amount reflecting the value of the resources used in making it. When using a commercially provided service, there will obviously be a cost involved to the user, namely the price or fare charged. But for an efficient outcome, this price must also reflect the value of all the resources used in providing the service that could be used to produce something else of value.

In the case of intercity traffic by highway, the issue of highway pricing or user-charges is of critical importance. If, for example, heavy trucks pay less than the costs they incur or impose, more freight traffic will move by highway than rail, other things being equal, which will create more highway congestion. This will raise costs for all highway users including cars, trucks and buses and will reduce quality (e.g. slower speed, more accidents, pollution, etc.). More accurate cost-based pricing for highway use might shift some traffic back to railroads and reduce highway congestion and pollution along corridors heavily used by trucks not paying their marginal social

costs. Clearly, a viable intercity commercial passenger bus industry as well as private automobile users have a large stake in an efficient system of highway user-charges. The railway industry has historically complained about federal and state subsidies to trucks and buses by virtue of under-charging them for the costs of using the publicly provided rights-of-way and thus creating a competitive disadvantage. For the sake of efficiency within the entire transportation industry, both passenger and freight, a set of at least reasonably plausible, cost-occasioned user-charges is indispensable.

A properly designed fuel tax supported by licence and registration fees for vehicles, trailers and other highway user types, can come tolerably close to what is meant by marginal social cost. Indeed, this tax itself is a kind of synthesized price for highway services. It is not of course easy to determine the economically correct level of taxation nor the correct value of marginal cost for each segment of the overall highway system. There are many conceptual, empirical and analytical problems involved. But acceptance of the principle of linking specific taxes to variable highway costs will provide incentives to improve and sustain such a linkage especially in the context of a separate trust fund. It is in this sense that the creation of the Highway Trust Fund in 1956 has such an important set of implications for any policy that places greater reliance upon market forces. Highways themselves are productive assets and should be priced appropriately. The U.S. highway system represents a huge investment, and its use, expansion and improvement needs to be made on economic principles as long as we live in a world of relative scarcity because the resources used up on highways have alternative uses whose value exceeds zero.

CONCLUSION

The post-war period through the 1950s saw significant further improvements in air and road infrastructure and a large expansion in use. The principles of more economical finance at least of highways were recognized and set into law. One could even see the glimmer of emergence of market orientation in economic regulation but not much. Yet in the background was a growing agitation among academics to change the system. Many others were becoming alarmed at the periodic crises that seemed to arise in all modes of transportation but mostly in rail. Debates were begun about what to do about it. Transportation, in short, began to rise in the national priority list of areas that needed attention.

5. EVOLUTION OF PRO-COMPETITIVE POLICIES: 1960–1982

The two decades beginning in 1960 saw the increasing strength of the regulatory reform movement and their ultimate achievements. These consisted of complete success in eliminating economic regulation in the airline industry along with the CAB itself, substantial deregulation of railroad, truck and waterway freight transport, elimination of most restrictions on bus passenger traffic and the removal of rail passenger service from the railroads themselves and integration of the service into a single, government-created organization called the National Railroad Passenger Corporation (Amtrak). Amtrak was given much discretion over fares and service but was also heavily subsidized and thus indirectly “controlled” by the government. The ICC remained in existence with sharply curtailed functions and authority although it managed to keep busy with aspects of the freight business that on the surface were surprising. Thus for example, the *Staggers Rail Act* that “deregulated” rail freight invited the carriers to discriminate in rates in a fashion that had led to the original Act in 1887 in the first place. Discrimination was supposed to be necessary to make the carriers “revenue adequate” and initially was constrained within a range of certain percentages above out-of-pocket cost (or marginal cost) for particular traffic. The Commission was to determine when carriers were “revenue adequate,” what constituted marginal cost and make other findings such as “market dominance” as well. In short, the ICC was to preside over a government-mandated system of discrimination previously declared illegal. Clearly all was not as many economists had envisioned, and parts of the regulatory reform “victory” seemed hollow indeed. But the main point was that efficiency criteria were given primacy in most cases and that individual carrier discretion was vastly enlarged. Even discriminatory freight rates could be viewed as efficient when carriers were subject to an overall revenue constraint. A theorem, now called “Ramsey pricing,” first developed by the eminent economist Frank Ramsey in 1928 and widely ignored by those economists unable to see anything beyond the marginal-cost-equals-price condition for efficiency, proved that for multi-product firms subject to a revenue constraint, prices which deviate systematically above marginal cost on the basis of the inverse of the elasticity of demand are optimal in the consumer surplus sense. Even the ICC pretended to espouse Ramsey pricing in a series of early cases after 1980 but, of course, could not do it correctly for reasons that need not be pursued here.

Regulatory reform of passenger business was virtually complete. The airlines, which carried almost 84 percent of commercial intercity passenger miles in 1980, were provided with most economic freedom which they had fought tooth and nail for over a decade. Passenger buses, accounting for about 11.5 percent of passenger-miles, were significantly deregulated while a single-firm monopoly, Amtrak, generated the remaining trivial amount. Thus well over 90 percent of intercity passenger transportation in the U.S. was, by the early 1980s, influenced almost totally by so-called "free market forces" in most of the relevant submarkets, namely, city-pairs. Certainly the goal of greater reliance upon market forces had its fullest expression in passenger transportation as compared with freight.

However it took over 20 years in the U.S. after the beginning of strong attacks upon the then existing commission-type regulation to bear fruit despite the ardent support of every president since Eisenhower. Even Eisenhower was sufficiently upset with the existing arrangements in transport to have called forth the Weeks Report which noted the end of the previous transport monopoly since 1920, talked about obsolete regulation and urged "Increased reliance on competitive forces of transportation in ratemaking."⁷⁸

Economists, political scientists and other academics and the legal profession had likewise begun the attack in the 1950s and some much earlier. The procedures of regulatory commissions, in transportation and elsewhere, were severely criticized leading to passage of the *Administrative Procedure Act* (1946) which specified the requirements for fair hearings for all affected parties — a set of rules that was so abused by various regulatory commissions, including the CAB and to a lesser extent the ICC, that docketed cases took an average of almost three years to decide and many important cases took far longer. Regulatory delay became a basis for advocating reform long before 1960.

Before 1960 many observers had noted the decline in the apparent quality of regulatory Commissioners. No new Joseph B. Eastman arose in the ICC and the last CAB chairman before Alfred E. Kahn was described as the "worst ever." Not only were the Commissioners of poorer quality and often "political appointees" in the worst sense of that term, but their frequently cozy relationships with the presidents of the carriers regulated led to the widely held "capture" theory and questioning of just how "expert" and "independent" the Commissioners were.⁷⁹

Thus the attacks upon regulatory commissions began much earlier than 1960. But the *nature* of the attacks shifted from procedures, internal operations and organization, venality and quality of the commission and its staff and the like, to the *effects* of whatever the regulatory agencies did upon the industries they were supposed to be regulating in the public interest. At this point, a bevy (perhaps “pretension”?) of economists turned their attentions to transportation once again as they had prior to 1930. As noted in Section I, the economic consequences, when properly advanced both in theoretical but especially in empirical terms, caught the attention of Congress and all the presidents since Kennedy. The present section outlines this story and seeks to explain both how anti-regulation forces finally succeeded as well as they did and why they took so long.

We begin with one of a series of reports and messages that tackled the problems of U.S. transportation and sought to provide solutions that would lead to sensible change. The change could take several forms: it could be new legislation mandating specific changes in the basic regulatory statutes and/or it could be changes by the commissions themselves in interpretation of existing law. For example, some years before legal changes were passed by Congress, both the ICC and CAB began to significantly alter past practices mostly in response to the ground swell of opposition to the *status quo* but also to avoid new Congressional mandates. It was said that deregulation of trucking occurred almost four years before the *Motor Carrier Act of 1980* because the ICC allowed almost every application for new entry into hitherto heavily protected markets starting in 1975-76. The CAB began urging new rate initiatives and permitted easier entry even before Chairman Kahn arrived on the scene which was several years before Congress passed the actual deregulation bill. Discussion of which approach is preferable will be deferred until the regulatory reform story has been examined.

THE DOYLE REPORT⁸⁰

This report, appearing in June 1961, shortly after the Landis Report had castigated regulatory agencies in general (see below), gave a further set of criticisms directed at the regulation of transportation, arguing that the “framework of present regulatory policy . . . has produced a *general program of preserving the status quo which is in direct opposition to the overall objective of a dynamic transportation system which can best serve the economy and defense of the country* [emphasis added].”⁸¹ The Report

then emphasized the essentiality of preserving common carrier service as the backbone of the system and stressed the need for regulation because "transportation partakes of the dedicated nature of public utilities. . . . [It is] in the public interest that a degree of stability and uniformity be introduced in the rate structures of the several modes . . . that at no time in the course of our review . . . have we found any serious recommendation that specialized regulation of transportation be discarded."⁸² In particular the Report stressed that intermodal cooperation is desirable but that it cannot come about through voluntary means because there are "too many carriers . . . each promoting his own immediate self-interest . . . too much mutual distrust and antagonism."⁸³ Thus the desired amount of coordination "will have to come about through regulation or as a result of permitting ownership of one mode by another."⁸⁴ Even though the latter is to be encouraged and the Report so recommends, regulation is still needed.

On the general issue of deregulation and pro-competition, the Report takes a large step backwards. Thus, for example, it argues that "With the exception of farm to first market and certain possibly noncompetitive traffic . . . exemption from regulation of for-hire transportation is contrary to national long-range interest."⁸⁵ Indeed it seeks through appropriate regulation to achieve "the right amount of competition" and that this will "minimize the cost of transportation to our economy"⁸⁶ and lead to rates that reflect long-run marginal cost. Competition, on the other hand, will not "tend to produce rates closely related to long run marginal costs" and there are sound and "logical reasons for concluding that competition in transportation rate-making will not automatically tend to bring rates and costs into a close relationship."⁸⁷ Such relationships require a rate *policy*; that is, rate regulation. These kinds of statements buttressed by much argument in an otherwise rather thorough and well-conceived report, help explain why the movement toward policies stressing "greater reliance upon competitive forces" became so protracted in the U.S. Its study director was a rather well-known and prominent economist who should have known better.

The Report, while a little weak on the economics of inter and intra-modal rate competition, nevertheless is all-encompassing. After careful analysis of the specific issues, it made such recommendations as:

- more careful assessments of user charges for the already existing highways;

- the creation of an airway trust fund which is now in existence;
- creation of a Department of Transportation *mainly to coordinate public transportation investments* with apparent need to prevent excess capacity;
- development of a waterway trust fund and full user-charges for use of the publicly improved waterways despite contentions that they were to be “forever free,” a contention which the Doyle Report labels erroneous and misleading;
- a careful assessment of urban transit problems;
- encouragement of transportation companies, as noted above;
- enlargement of motor carrier operating rights and so on.

Labour problems in transportation are discussed which usually are ignored. Many other detailed analyses of aspects of current policies that it sees as causing malaise for common carriers of freight are included.

However, in addition to urban transit problems, the Report also addresses intercity passenger transportation which is more directly germane to this essay. It notes that “Railroad intercity passenger service meets no important needs that cannot be provided for by other carriers, . . . possesses no uniquely necessary service advantages . . . [and] serves no locations which cannot be served by air or highway;”⁸⁸ it can only earn a place by “offering a combination of price and convenience which will attract business in open competition at full-cost fares.”⁸⁹ The Report believes that so oriented and operated only in high-volume markets, some intercity passenger traffic could survive or at least “produce a far higher ratio of revenues to expenses.”⁹⁰ The causes of the demise of this service include rail management apathy and unwillingness to pool equipment and services to reduce duplicate facilities. This in turn is related to the lack of “effective profit-responsible management of this service.”⁹¹

To be sure, regulation and government creation of excess air and highway facilities have not helped. Yet the Report feels that all is not necessarily lost. “If this service goes to the museum with the stagecoach [as the Hosmer Report predicted], it will not be for the same reasons.”⁹² If existing management is incapable of overcoming “the problems of equipment interchange, trackage rights, financial realignments, managerial realignments and eliciting

the reasonably unified cooperation of over 80 companies, *a national railroad passenger service corporation should be considered* [emphasis added]."⁹³ This more or less prescient proposal was however not to be implemented unless a recommended study indicated potential profitable operations based on conservative traffic and financial forecasts.⁹⁴

This Report, coupled with the Landis Report to President-elect Kennedy, led to some administrative and procedural changes within the ICC and CAB which amounted to very little. Nor did the Report lead to new legislation possibly because its coverage of topics was so extensive, some of its recommendations exceptionally offensive to trucking and waterway interests, and parts of the analysis purporting to justify continued rate regulation were out of step with the times as well as analytically slipshod.

However, it represented something of an advance over the Weeks Report, continued the attack on the transportation regulatory agencies and reinforced the need for change. It was also a creature of the U.S. Senate and indicated a growing concern in that body for regulatory reform.

The Landis Report, noted above and in Section I, although not focussing upon the transportation agencies exclusively, kept the heat on all the agencies and indicated serious concern from the White House as well as Congress. Specifically, James Landis, who once developed the rationale for a compelling need for regulatory agencies, the "fourth branch of government," in the 1930s, now engaged in a merciless dissection of the agencies' failures, possibly out of a sense of betrayal of his regulatory ideal.⁹⁵ Emphasizing the appointment of weak, vacillating and non-expert commissioners during the Truman and Eisenhower years, the problems of capture and especially delay, the absence of any general policies for particular problems combined with the multiple impacts the agencies felt they had to consider,⁹⁶ the absence of concern for efficiency and so on were all highlighted.

Regulatory reform began to shape up as a leading issue in the first year of the Kennedy administration. While no new legislation was passed, the appointment of superior and highly motivated people to such agencies as the SEC, FCC and FPC led to intra-agency shakeups along similar lines that Alfred Kahn was to follow in deregulating the airlines. If Landis could complain in his report to Kennedy that "the fires that then [1930s] fed a passion for public service have burned low," Kennedy helped reignite them in the

early 1960s. But alas, it was far more than inadequate personnel that ailed economic regulation of transport — it was the entire foundation and rationale for regulation and its increasingly costly consequences that needed change.

PRESIDENT KENNEDY'S MESSAGE ON TRANSPORTATION⁹⁷

The assault on transportation regulation continued when, by inadvertence, President Kennedy *ad libbed* during a speech that he was going to deliver a message to Congress on transportation in the near future. Nothing of the sort had been previously discussed.

Released on April 5, 1964, the message was “Hailed as the most comprehensive transportation proposal a president ever submitted to Congress.”⁹⁸ Written by economists both within and outside the government, it bristled with concern for efficiency, low cost, economical service and so on. Noting that “pressing problems are burdening our national transportation system” and that existing regulations are a “chaotic patchwork,” Kennedy argued that “less Federal regulation and subsidization is in the long run a prime prerequisite of a healthy intercity transportation network.” The national policy required, among other things, that “the resources devoted to provision of transportation service should be used in the most effective and efficient manner possible; and this, in turn, means that users of transport facilities should be provided with incentives to use whatever form of transportation . . . provides them with the service they desire at the lowest total cost, both public and private.”⁹⁹ This objective required unsubsidized, privately-owned facilities operating under the checks of competition to the maximum extent possible and reduced regulation. Users should also bear the full costs of both publicly and privately provided facilities. To move toward these goals, he recommended extension of certain commodities now exempt from all rate regulation for water carriers (bulk commodity exemption) and for motor carriers (agricultural and fishery products), to all carriers. Congress was invited to enact legislation to *limit the control of intercity passenger rates to the establishment of maximum rates only*.¹⁰⁰ And so it went on to recommend “consistent policies of taxation and user charges” even for waterways; even-handed government promotion of intercity transportation mostly by phasing out subsidies; careful scrutiny of large mergers to maintain as many shipper and traveller options as was efficiently feasible; encouragement for establishing through routes and joint rates; and general stimulation of

experimental rates, fares and services. Beyond intercity transportation, there is a long section on urban transit, one of the first of its kind and a major section on international transportation — neither of which are germane for present purposes but which do indicate the scope of the message.

It received wide press coverage and was greeted with much enthusiasm. However nothing came of it in terms of legislation. It lit a few more fires for dedicated public service and furthered the cause of transportation reform but it evoked substantial and powerful opposition. For example, a bill proposing exemption from minimum rate regulation for intercity passengers and transportation of bulk commodities and fishery products was opposed by the ICC, the American Trucking Associations, the International Brotherhood of Teamsters, the Association of Motor Bus Operators, the various waterway associations and many other groups. The grounds for opposition were mainly that equalization among modes should move in the direction of *greater* regulation of all of them, not less. The Association of American Railroads, the U.S. Chamber of Commerce and various agricultural, coal and shipper groups supported the proposals but only if they were amended to retain the exemption for the railroads in their minimum rate-making from the operations of the antitrust laws, an amendment rejected by all the other groups noted above who opposed the bill in the first place.

Legislation designed to implement other parts of the message evoked similar opposition. None ever got out of committee. The anti-deregulation forces were more powerful than recognized up to that time. There was widespread belief, especially within transportation circles, in the ICC's statement in opposition to the bill:

If transportation teaches any one thing, it is that while competitive forces generally are effective in reducing prices and improving standards of service, these very same competitive forces in the transportation field, unless subject to reasonable restraints, will result in eliminating competition and in disrupting reasonable and fair rate relations as between competing shippers, geographical areas and territories.¹⁰¹

This reflects and reinforces the view that in some sense transportation is unique, that it is vitally important, has enormous positive and negative externalities, and so on that require regulation of some kind beyond the

discipline of competitive markets. Even the Doyle Report had some of this attitude. Yet the sense in which it is unique is seldom spelled out. Certainly the cost functions of the various modes do not differ from those of most other industries. The fact that transport provides a perishable service is scarcely grounds for alarm in an economy where two thirds of GNP is generated by services rather than goods production. Nor are any of the modes "natural monopolies" in the sense of being subject to scale economies. Obviously air, motor, bus and waterway transport of people or goods are far from being naturally monopolistic. There is less assurance about railways, but for many decades the view that unit costs decline with volume of production of either passenger- or ton-miles was regarded as demonstrating economies of scale when in fact it represents economies of density — quite a different thing. Railroads are probably *not* subject to economies of scale. Even if they are, there are upper limits to any rate or fare provided by other modes so long as all markets are reasonably contestable by any company that is "fit, willing and able."

Sometimes the "uniqueness" theorem is supported with the observation that transportation is essential for all societies for a whole lot of reasons that are too self-evident even to mention. However, essentiality is not unique. Food is essential for life and food production in most economies is usually left to market forces or, if not, should be. Indeed, agriculture is viewed as the quintessential industry where market forces work best if governments refrain from mucking them up through price supports, acreage limitations and other intrusions. The same is true of health care, housing and other so-called necessities of life. None of them have been subjected to the same degree of protective, promotive and detailed intrusion into every aspect of the businesses providing them as transportation, especially in a free enterprise economy.

The economic virtues of private ownership, operation and decision making in response to profit opportunities, suitably constrained by competition and circumscribed by general rules of the game set down in law and subject to change via democratic processes, has been apparently vindicated by worldwide events of the last two years. Most assuredly this does not mean *laissez-faire* in the extreme sense. Even Adam Smith recognized legitimate and necessary roles and activities for governments. Some of these issues involving transportation will be discussed in Section 6. But some obvious ones may be mentioned here. They include antitrust laws to preserve

competitive options; efforts to eliminate negative externalities consistent with market processes such as making private and social costs coincide as closely as possible; and efforts to maintain full employment, steady growth and a tolerable income distribution. Changed conditions in transportation had created, long before the end of World War II, a set of circumstances which rendered the pre-existing regulations of the industry at best inappropriate and at worst far more costly than any discernible net benefits to society and the economy as a whole. This was the message that was left to the economists to present as forcefully as possible. Part of this required debunking the uniqueness theorem and with it the Grand Transportation Mystique.

The strength of the opposition to regulatory reform seemed rather surprising at the time (early 1960s) as did the failure of the Kennedy message, given his initial enormous popularity. What was not recognized was the extent of the vested interests that had grown up around the existing regulations. In addition to heavy Congressional lobbying by railroads over many years, joined by the increasingly powerful American Trucking Associations, the Teamsters Union and others noted above, there were considerable financial contributions to the campaigns for electoral office at all levels of government by the various modal and other interests concerned with keeping things as they were. This is, of course, standard operating procedure in the U.S. even though there have been constant attempts to control its abuses over the years. Nonetheless, it remains a powerful force for maintenance of the *status quo*. But in addition, the laws affecting transportation economic regulation were so complicated and subject to so many conflicting interpretations that an entire cadre of transportation lawyers for decades has made a very good living from representing transport firms before the ICC, CAB and the courts at all levels. With them, accountants, statisticians and economists were increasingly involved. Hundreds of private consulting firms have found their bread and butter in transport cases, litigation and simple hearings before administrative law judges.

Even when Congress thought it had changed the law, as in the *Transportation Act of 1958*, the ICC interpreted it in such a way as to imply that little had really changed. This was also true of parts of the so-called 3R and 4R Acts passed in the 1970s. The Commission's scope for interpretation was extremely broad and it used this to maintain the *status quo*. However, this had one possible advantage, namely that if new, reform-type Commissioners could be appointed, sensible changes might be feasible without new

legislation which took so long, required so much compromise, and even then was uncertain. Indeed, I once argued, prior to the legislation of 1980, that "there is nothing in the *Act to Regulate Commerce* as it now exists to prevent the ICC from doing any or all of the things that the proponents of regulatory reform recommend. The reason we have been forced to legislate is that the commission will not do them . . ." ¹⁰² If only it had been this simple! As many have pointed out, for any long-range guarantee that the reforms will stick, legislation is needed for the very reason it is harder to obtain in the first place. One can replace Commissioners with different viewpoints even if it takes a while (that is, at most one term if resignation cannot be induced or ill health speed things up) so that new administrations can easily re-regulate. Indeed, there is now a growing pressure for re-regulation in the transport industries in the U.S.

AFTER PRESIDENT KENNEDY

After President Kennedy's death in 1963, President Johnson accepted deregulation and the emphasis upon efficiency in transportation. While declaring the high importance of transportation reform he soon became obsessed with the civil rights movement and increasingly with the Indochina War. Unable to devote much time to transport, he nevertheless succeeded in creating the Department of Transportation (DOT), a long-time objective of many reformists and advocates of more coordination, which is to say, almost everybody. Thus, although sending a transportation message to Congress that emphasized the need to create a National Transportation Safety Board, to do more research on the SST and high-speed ground transport, and promoted transport coordination through a new federal department, only widespread, general support was forthcoming for the DOT, which began operations on April 1, 1967. Although not in direct opposition to the ICC, the DOT has supported regulatory reform and has made legislative proposals usually supported by the White House almost every year. Most got short shrift until the late 1970s.

Johnson's Council of Economic Advisers began a series of what were to become increasingly strident criticisms of transport regulation in each of its Economic Reports. Beginning in 1966, a section entitled "Efficiency in Transportation" was included. It advocated the DOT, of course, and discussed the role of competition in forcing rates toward marginal costs. But it was still partly in the regulatory mold. "While controls over entry and abandonment

are surely *desirable* [emphasis added]," it opined, "considerably more flexibility would seem to be appropriate."¹⁰³ This is pretty timid stuff but the beginning of another aspect of the attack on the transport *status quo* that continued fairly regularly and broadened in scope through the 1990 *Economic Report of the President*. What clout it had is difficult to say. In the 1960s, the Council of Economic Advisers had considerable stature, status and support of the President but its repute had dissipated by the 1980s. It was, however, only one facet of the economists' continuing attack on transport regulation. The attack was rendered far more powerful as economists were hired in ever larger numbers and quality in various branches of government.

THE ECONOMISTS' ROLE IN TRANSPORTATION REFORM

The 1960s ended without much in the way of regulatory reform. However, the economic case for it had been well prepared and increasingly documented since the late 1950s and throughout the 1960s. In very general and overly brief terms, the case involves two interrelated parts. The first is that since the 1920s at least, the technical changes in transport — namely the automobile, truck, bus and airplane, and the associated highway and airway investments and progressive improvements in all phases of these technologies — had created an enormous and growing supply of transport capacity alongside the earlier dominant railroads already suffering from overcapacity. Thus the potential competitiveness of transportation between most city-pairs was vastly increased. Virtually all such markets had a sufficient number of traveller and shipper options supplied by independent carriers to make competition at least workably competitive. At the same time, it made economic regulation incredibly complex if not virtually impossible. In short, regulation became impossible to perform with any degree of rationality, reasonableness or efficiency; at about the same time, give or take a decade or two, it became unnecessary.

The problems of regulating the "new" industries of highway transport (truck and bus) and airlines which used publicly provided, improved and maintained rights of way, with little evidence that particular users paid appropriate amounts for either the costs they occasioned by such use or benefits received, created sharp differences among the modes including the railways. Any attempt to regulate for competitive equity or even to discover, let alone preserve "inherent advantages," as the ICC was enjoined to do, was virtually impossible. The large number of firms brought under regulatory purview,

especially that of the ICC, with the necessity of making “grandfather” exemptions, exemptions for private carriers in motor transport and lesser regulation of contract carriage by truck or bus, vastly complicated the regulatory process. Meaningful distinctions among common, contract, exempt and private carriage are impossible to make and enforce. When certain commodities were declared exempt by statute, the situation became even more absurd. Court cases and much Commission time was expended on such trivia as whether frozen chicken is a manufactured or an agricultural, and hence exempt, commodity.

All the regulatory agencies became weighted down with trivia. Cases took years to conclude in part because of the new complexities but also because of the over-zealous use of “due process.” The vastly different cost and demand functions relating to these disparate industries defied anything like a unified, basically anti-competitive approach that attempted to preserve something referred to as “common carriage,” itself becoming increasingly obsolete and irrelevant with the new powers of own-account transport of people and goods. The ICC even used different profit concepts for rail and truck or bus, namely rate of return for the former and operating ratio for the latter. Even with enormous wisdom and administrative talent, the Commissioners could not cope with such a morass especially when the legislation they were operating under was vague, subject to a wide range of interpretation and in many respects inconsistent. But during the 1940s and 1950s and even later, Commissioners were appointed more frequently on political grounds and were distinctly inferior in their concern for and knowledge of the industries under their charge.

Given these considerations, the ability to carry out any grand Benthamite design was rendered worse than impossible. There was no way the Commission could possibly weigh the advantages and disadvantages to shippers, travellers, regions, ports, national defence, the needs of commerce and the Post Office in any given decision. Not only could the benefits and costs not be measured with any degree of accuracy but there was no weighting system that made any sense nor did the Commission ever reveal any systematic set of preferences except for common carriers. In fact, the weights seemed to vary from case to case. Thus the system collapsed for the same reasons that Bentham’s felicific calculus collapsed — an inability to measure or weight the impacts using any standard metric that was additive. The Commission’s decisions then became whimsical, without a

rationale that could even be used as a precedent in other cases because the relative importance of the several impacts upon the multiple considerations alleged to be considered varied from case to case. This was *sui generis* with a vengeance.

Although the CAB never adopted an apparent grand design of such scope, it ran into similar problems because it sought to create and administer a cartel, oriented to sustaining the profitability of the *scheduled* carriers over time. If the economy was buoyant and travel demand brisk, the Board could be more lenient in granting new services or increased competition on certain routes and even encouraged discount fares. On the other hand, when times were difficult and carrier profits threatened, the Board allowed few new entrants and discouraged price reductions contrary to what a market-oriented policy would require. Thus the policy seemed to fluctuate between pro- and anti-competitiveness depending upon the general economic situation. As has been noted, "The CAB usually was vigorously anticompetitive when the airlines were suffering financially, and it was relatively procompetitive when the airlines were prospering. Thus . . . CAB policies between 1938 and 1974 were a result of applying stable protectionist principles to changing economic conditions."¹⁰⁴

Thus the Board could look as if it were not so flagrantly against competition from time to time even though it thoroughly applied controls over entry and competition for the benefit of a favoured few — the scheduled carriers. It was thus more able to carry out a more limited goal of seeking stable profits than the multiple goals of the ICC. Yet even with a single goal it was not very successful in creating stable earnings year after year and certainly not for each carrier. However, there were no bankruptcies among the majors. The airlines themselves favoured continued regulation, as did airline labour, among the highest paid in the country, and financiers of the industry. Despite the economic arguments about misallocation, excessive prices and costs, lack of price-quality options and other problems, there was little enthusiasm for deregulation of the airlines or for reducing significantly the regulatory powers of the CAB. Despite this, potential competition was constantly growing.

In general, regulation of transportation became increasingly complex, less effectively done or doable and less necessary as the competitive potential of transport markets grew. The "costs" of regulation began to exceed the "benefits" presumed to derive from it. Even the railroads came to support some deregulation by the late 1970s as their economic situation worsened.

Legislators heeded the deregulation cry when various economists began to calculate the costs of regulation. Such costs turned out to be enormous, between \$8 and \$16 billion in 1977.¹⁰⁵ Though based upon very flimsy assumptions which no one chose to question, they nevertheless caught the attention of legislators far more directly than generalized statements about the competitiveness of most transport markets and the supreme virtue and efficiency of market solutions over regulatory ones.

The task now became how to get these virtually unanimous conclusions about the need for transport regulation reform, something very rare among economists, out of academia and the professional journals and into the mainstream of policy discussion and decision making in Washington. This process was aided by several Washington "think tanks" including the Brookings Institution, probably the best of them all, and one or two others which sponsored programs of studies on economic regulation of business and produced numerous volumes, articles, pamphlets and conferences on the findings. Most of these were written or conducted by scholars in economics, political science and other policy-based, social science research. This helped capture attention in the right places for policy reform in transportation and gave experience to academics who would be appointed to government agencies also concerned with the issues.

Such agencies as the Council of Economic Advisers, the Antitrust Division of the Department of Justice, the Department of Transportation (after 1967), the Office of Management and Budget as well as presidential task forces and Congressional studies (such as the Doyle Report and the National Transportation Policy Study Commission of 1979) virtually gobbled up economists who evinced any talent and/or interest in the economics of reform especially if it were pro-competitive. Since the profession as a whole was pretty unanimous on this point, there were many economists available. As has been said, by "1970-71, there were enough active advocates of regulatory reform inside the government to constitute a small, informal community, loosely bound by personal acquaintance and commitment to a shared policy goal. What had begun as a random critique of government policies . . . in a decade had evolved into a moderately concerted effort with a substantial official and quasi-official base."¹⁰⁶

Economists were now in a position to do something more than simply argue the case. At the centre of much of this activity was transportation since it was not only a large and important segment of the economy but the first to

have been subject to the kind of economic regulation now under attack and in which regulation had been hastily expanded to include railroads' competition thereby providing further ammunition for reform.

Almost from the beginning, the reform movement in transportation had support from the executive branch. Presidents Kennedy, Ford, Carter and Reagan gave high priority to it as did Johnson and Nixon, although the latter were unable to spend much time or effort in the cause.

However, to obtain the desired legislative change required sympathetic understanding from Congress. For a variety of reasons, the issue of transport reform, especially of air and truck transportation, was taken up in the early 1970s by the most publicized Democrat and one of the most powerful Senators in the legislative branch, Edward M. Kennedy. His advocacy virtually ensured passage of the bills deregulating both airlines and motor carriers. The story of airline deregulation and Kennedy's role in it will be told later in this section.

One more ingredient was needed to get legislation passed. It had to be shown that without it there would be dire consequences for the nation. Transportation had to become a high-priority item on the national policy agenda. However, the 1960s did not indicate much of a crisis. Rail passenger deficits remained high but stayed below their peak levels (which averaged \$680 million in the seven years before the *Transportation Act of 1958*) averaging about \$420 million per year from 1960 to 1970 as the ICC accelerated abandonment. (See Table 2.) There seemed to be no railroad "crisis." Airlines were doing quite well, buses continued their unspectacular but not dismal performance and the economy was progressing nicely with low unemployment rates, and better than average economic growth. Although fears of inflation led to a tax surcharge in 1968 and the Vietnam War had occasioned urban riots, campus uprisings and a souring, contentious mood throughout the country, transport problems, when thought about at all, were far from uppermost in public perception. Invoking the admonition "If it ain't broke, don't fix it," there seemed little point in using up political or other capital to tackle the problems the academics saw in transport. "Thus procompetitive regulatory reform was well and widely articulated as a policy prescription; but it remained a solution in search of a widely perceived problem . . ." ¹⁰⁷

Table 2

OUTPUT AND FINANCIAL PERFORMANCE OF AMERICAN RAILROAD PASSENGER SERVICE, 1920-1970

Year	Passengers (thousands)	Passenger-miles (thousands)	Net revenue ^a (thousands of dollars)	Passenger deficit as percentage of freight net revenue
1920	1,269,913	47,369,906		
1921	1,061,131	37,705,737		
1922	989,509	35,811,046		
1923	1,008,538	38,294,178		
1924	950,459	36,368,290		
1925	901,963	36,166,973		
1926	874,589	35,672,729		
1927	840,030	33,797,754		
1928	798,476	31,717,566		
1929	786,432	31,164,739		
1930	707,987	26,875,642		
1931	599,227	21,933,345		
1932	480,718	16,997,426		
1933	434,848	16,368,635		
1934	452,176	18,068,635		
1935	488,059	18,509,497		
1936	492,493	22,459,781	-233,327	26.2
1937	499,688	24,695,214	-241,591	29.2
1938	454,508	21,656,918	-255,263	40.8
1939	454,032	22,712,941	-250,934	29.9
1940	456,088	23,815,598	-262,058	27.8
1941	488,668	29,406,250	-226,029	18.5
1942	672,420	53,747,029	89,329	—
1943	887,694	87,924,994	279,790	—
1944	915,817	95,662,501	234,103	—
1945	897,384	91,826,353	230,060	—
1946	794,824	64,753,699	-139,736	18.4
1947	706,551	45,972,245	-426,526	35.4
1948	645,535	41,224,319	-559,782	35.9
1949	556,741	35,133,300	-649,627	48.6
1950	488,019	31,790,470	-508,508	32.9
1951	485,468	34,640,031	-680,822	41.9
1952	470,979	34,033,245	-642,390	37.3
1953	458,252	31,678,951	-705,538	38.9
1954	440,770	29,309,861	-669,533	43.4
1955	433,308	28,547,877	-636,693	36.1
1956	429,994	28,215,728	-696,938	39.5
1957	412,625	25,914,446	-723,483	44.0
1958	381,623	23,295,262	-591,543	35.7
1959	353,647	22,074,718	-523,692	32.8
1960	327,172	21,284,084	-466,289	32.9
1961	318,359	20,308,444	-390,495	29.7
1962	313,084	19,926,466	-374,993	25.2
1963	310,999	18,519,049	-378,618	23.7

Table 2 (cont'd)

OUTPUT AND FINANCIAL PERFORMANCE OF AMERICAN RAILROAD PASSENGER SERVICE, 1920-1970

Year	Passengers (thousands)	Passenger-miles (thousands)	Net revenue ^a (thousands of dollars)	Passenger deficit as percentage of freight net revenue
1964	314,386	18,271,322	-389,008	23.7
1965	305,822	17,161,776	-398,029	21.6
1966	307,530	17,162,776	-379,744	19.5
1967	304,028	15,264,172	-460,414	26.9
1968	301,372	13,163,541	-462,129	25.8
1969	301,673	12,213,983	-437,498	24.4
1970	289,469	10,785,746	-449,579	26.2

Sources: Interstate Commerce Commission, *Statistics of Railways in the United States; Transport Statistics in the United States*; James C. Nelson, *Railroad Transportation and Public Policy* (Washington, D.C.: The Brookings Institution, 1959).

a The ICC did not separate passenger and freight net earnings before 1936.

The crisis to advance regulatory reform was not long in coming. Indeed in three successive crises or at least serious and widely perceived events, transportation in general and railroads in particular were catapulted into the limelight between 1970 and 1973 (a light they shared with Watergate and the 1972 elections). The first of these was the Penn Central financial collapse in 1970, four years after the ICC had approved the merger of the New York Central and the Pennsylvania railroads and two years after the Supreme Court had finally sanctioned it. The Commission's approval was an obvious mistake not only for commercial reasons, as the bankruptcy shortly thereafter confirmed, but for reasons of sensible public policy. For example, as I wrote to a colleague shortly after ICC approval in 1966 under the heading, "Reflections on the N.Y.C.-Pennsy Merger";

The approval of this merger by the ICC clearly flies in the face of predominant thinking with respect to deregulation, rate freedom, and increasing reliance on competitive forces in transportation. The foreclosure of intra-railroad competition over the vast area encompassed by such a merger should give those advocating reduced regulation occasion for some second thoughts. Whatever one may think of the possibilities of intra-railroad competition and its potential effectiveness, it cannot be lightly discarded as a totally inconsequential force. The

entire philosophy of increased reliance upon competitive forces presupposes the existence of competitors having roughly similar cost and service characteristics. Intermodal competition, to be sure, is vital and in general ubiquitous within regions but to increase competitive effectiveness, intramodal competition is also essential. Thus the issues of rate freedom, deregulation and mergers must be viewed as a whole. Cartelization of any large area of potential competition requires more effective regulation of rates, fares and service, not less. Thus, the failure of the Commission to consider the broader picture in this instance jeopardizes any future move in the direction of deregulation. A cynic may even be led to infer that the Commission fully recognized this and approved the merger so that the necessity for more regulation, and with it the future life of the Commission itself, would be assured.

To be sure, perhaps the estimated \$80 million in savings will be achieved through elimination of duplicate facilities and all that. But against this must be set the dissavings that might be involved if the present regulatory system is thereby reinforced and rendered more impervious to change. I would guess that \$80 million is peanuts compared to potential savings through substantial deregulation.¹⁰⁸

Many people were incensed at the merger and frightened that the bankruptcy in 1970 would lead to liquidation. The combined roads constituted over 20 percent of total U.S. rail traffic and served a region of the country, the Northeast, that comprised 17 states and limited portions of other states, which accounted in 1970 for almost 60 percent of total industrial production and almost half the U.S. population. Any forced liquidation or even serious interruption of Penn Central service would have severe national impacts.

The second event, not unrelated to the Penn Central bankruptcy (possibly the most investigated and written about corporate failure in U.S. history) was the so-called Northeast Railroad Crisis. Referred to as "One of the most significant transportation occurrences of this century,"¹⁰⁹ it involved the bankruptcy of seven more railroads in the Northeast.

By this time the country, the Congress and the President were well aware of the railroad situation, its possible impact upon the nation as a whole and to a lesser extent how the causes were related to some more fundamental underlying malaise facing the entire transportation system. Out of these

problems arose several legislative actions, in 1970 alone. The *Emergency Rail Services Act of 1970* was a financial bailout, in the form of loan guarantees by the government (that is, the Secretary of Transportation) to permit operations of Penn Central and other financially strapped railways to continue. More important for present purposes, the *Rail Passenger Service Act of 1970* created the National Railroad Passenger Corporation (Amtrak), to remove the burden of losses from cash-flow poor railroads already in bankruptcy. The Penn Central alone is reported to have absorbed nearly one third of the total national rail passenger service loss — something it could no longer continue to do and which helped bring about its financial collapse in the first place.

There was clearly considerable fear and panic in the initial legislative responses to the crisis in the Northeast, especially the Penn Central situation. In fact another “crisis” situation, a one-day strike against the Penn Central and its bankruptcy court threatening a liquidation unless public funds were forthcoming, initiated a sequence of events that led to the *Regional Rail Reorganization Act of 1973*, the so-called 3R Act. Thus began a planning process that had been suggested as far back as the *Transportation Act of 1920* although now on a more regional scale. The earlier suggestion regarding planning had been ignored by the ICC. The results after two years of planning amid enormous infighting and litigation among government agencies, as well as involvement of the railroads and many other affected interests, were the establishment of the Consolidated Rail Corporation, Conrail, and an unprecedented legal situation by which the federal government had planned and implemented a mandatory restructuring of privately owned railroads. Cries of “unconstitutional” were widely expressed but later resolved. In all this, the ICC and rail regulation were ignored except for creating the Rail Services Planning Office (RSPO) as part of the Commission to help in the restructuring process.

In short, there was much crisis-oriented activity in attempts to respond sensibly to the twin crises of the railroads in the Northeast. The creation of two new quasi-government corporations, Amtrak and Conrail, the billions of dollars of bailouts (or, euphemistically, loan guarantees) or outright subsidies, obviously captured the public attention and clarified that there was much unfinished business with the U.S. transportation system as a whole including issues of regulatory reform.

But even these attention-getting events and their train of frenzied activity were upstaged by another more ominous occurrence in 1973. The invasion of Israel by Egypt and Syria on October 6, 1973 set off the so-called "Yom Kippur War." The results of this further evidence of implacable Arab-Israeli enmity induced a hitherto harmless cartel, Organization of Petroleum Exporting Countries (OPEC), composed of leading oil exporting nations, to ignore the private international cartel, composed of large petroleum corporations, and arbitrarily raise oil prices and embargo shipments of oil to the United States and The Netherlands, both active Israeli supporters. By January 1974 the price of oil was raised from about \$3 per barrel for Saudi Arabian light crude on October 17, 1973, to \$11.65 per barrel. This precipitated the first of the major supply-side shocks the world was to endure and come to fear for the next almost 20 years. It precipitated world-wide recession and was a turning point in world economic history as growth rates of all the then leading industrial nations slowed sharply and have yet to regain their pre-1973 levels.

But for our purposes the oil crisis, as it was often called, even though the world has abundant supplies of oil and is in no danger of "running out," focussed attention on the transportation industries. It was quickly calculated that in the early 1970s transportation accounted for over half of total domestic use of refined petroleum products and about one quarter of total U.S. energy consumption. Of this, automobiles and trucks accounted for about three quarters of transportation's energy consumption. U.S. imports of petroleum and products were less than \$4 billion in 1971 and earlier. By 1974 they totalled almost \$27 billion and increased to a high of almost \$80 billion in 1980. OPEC had seriously affected the world trading system and helped engender chronic balance of trade and current account deficits for the U.S.

This was serious business and called for an energy policy to stimulate greater efficiency in the use of oil and its products which in turn focussed attention upon transportation as a major user. To no one's surprise, it was calculated that railroads were the most efficient users of fuel for freight transport by a wide margin. BTUs consumed per revenue ton-mile of rail freight were far below those of truck or air. For passenger-miles the efficiency differences were less striking, but surprisingly Class I intercity bus travel was most efficient, followed by Amtrak, private automobiles and aviation the least efficient. See Table 3 for 1983 estimates.

Table 3

RELATIVE FUEL EFFICIENCIES: BTUs PER PASSENGER-MILE

Class I Intercity Bus	1,018
Amtrak	1,765
Automobile	3,498
Certificated Air Carriers	5,490
General Aviation	11,044

Source: U.S. Department of Transportation, *National Transportation Statistics* (Washington, D.C.: Government Printing Office, 1983), pp. 121-27.

These data, even if completely accurate, do not tell the whole story. They say nothing about the quality differences among the modes and are heavily influenced by the load factors on average for the several vehicle or plane types as well as variations among vehicles and aircraft themselves. Crude as they are however, they do suggest that the most fuel-effective way of moving people between cities is by rail and bus. Thus Amtrak and bus transportation should be encouraged relative to cars and planes especially general aviation! A new argument was added to go along with nostalgia and pollution to justify attempts to retain and stimulate rail and bus intercity passenger travel options and use, even if they needed to be subsidized. The excessive use of automobiles and air travel should correspondingly be discouraged. The value of time, comfort and convenience should be reduced and there should be efforts to improve the fuel efficiency of planes and cars. Attempts along all of these lines have of course been made and need not be detailed here. What is important for this discussion is how the above three events from 1970 to 1973 affected U.S. passenger policy during the 1970 to 1982 period when the last of the reform legislation was put in place with President Reagan signing the *Bus Regulatory Reform Act* in November 1982. The solution had finally found a widely perceived problem by the end of 1973. All the pieces were in place for regulatory reform.

The following discussion is confined to the three major changes pertaining to passenger transportation: (1) airline deregulation that went farther than any of the others; (2) the less discussed deregulation of the intercity bus industry; and (3) the deviation in the opposite direction in the creation of Amtrak.

AIRLINE DEREGULATION

From a position of almost zero interest in even modest pro-competitive reforms in 1969 to ardent advocacy in 1975 represents a sea-change for the CAB and the Congress.¹¹⁰ Of the several reasons for this rapid change, the first is rather paradoxical, namely an even more anti-competitive shift than before on the part of the CAB beginning in 1969, with the economic recession which carried over through 1970. The recession led to a sharp reduction in traffic at about the same time new wide-body jets, ordered several years earlier in anticipation of continuing high-traffic growth and route expansion, were delivered. The ongoing wage-price spiral inflated operating costs; financial costs were rising as well. The reduced profits and profit prospects induced the CAB to refuse all applications for new routes — a “route moratorium” as it was called. The Board also approved temporary capacity-limitation agreements among several large carriers to raise load factors, and further restricted scheduling and discount fare competition — a more perfect cartel was being formed in the face of a movement in the opposite direction. A new chairman even believed that these should be made permanent features of regulation!

The first oil crisis gave an excuse for fare increases and even more widespread use of capacity-limiting agreements (such as OPEC production quotas), ostensibly to save fuel. Profit margins recovered somewhat by mid-1974 about the same time the economy was falling into another recession, thanks largely to OPEC. The Board was thus seen as protecting one cartel while the country was suffering from the actions of another. This made it a prime target for change. Critics of regulation also benefitted from a minor scandal involving the Chairman of the Board who wanted to cartelize even further.

The hearings of Senator Kennedy began in February 1975 and provided a widely publicized forum for the economic case against continued regulation of the CAB type. The case was not especially abstruse and its critics had no substantive rebuttal. One of the more telling arguments showed fare differentials on intra-state, non-CAB-regulated carriers in California and Texas that were 50 to 60 percent lower than CAB-regulated fares for comparable distances in otherwise similar circumstances. The cut-rate fares had so raised the load factors for the intra-state carriers that it was highly profitable to charge them. The refusal to sanction rate competition had clearly denied the public the advantage of lower rates. As the Kennedy subcommittee

pointed out, Pacific Southwest Airlines "puts 158 seats in a Boeing 727-200 jet aircraft and fills approximately 60 percent of these seats on average. American Airlines puts 121 seats in the same plane and flies it on average 55 percent full. Thus, when flying PSA, 95 passengers must share the cost of flying the airplane, while, on an American plane, 66 passengers must share the same cost."¹¹¹ This was something anyone could understand and vividly indicated the benefits of more competition in rates.

In addition the hearings were sprinkled with the arguments that regulation leads to higher investment costs because carriers, unable to compete on price, seek the latest, fastest aircraft for competitive purposes and because of the A-J-W effect.¹¹² Higher operating costs occur because if the carriers cannot compete on the basis of price they compete on cost-enhancing amenities such as food, drink, movies, advertising, or gimmicks of one kind or another. In addition, other costs of opposing competitors' applications for competing routes or lower fares and the like as well as the trouble and expense involved in seeking permission from the regulatory body to do those things are not required in non-regulated or less-regulated markets. The loss of consumer choices regarding price-quality trade-offs that Freddy Laker was later to dramatize in transatlantic traffic, the extra costs of low load factors because of the failure to allow peak/off-peak pricing and so on were all trotted out and more or less "explained" to a public and a Congress suddenly aware that serious problems existed.

This was all reinforced by the powerful support provided by presidents Ford and Carter who made regulatory reform a major goal of their administrations. Airline deregulation had strong bi-partisan support.

Finally and not the least important were the chairmen of the Board, appointed by President Ford and President Carter. Ford's appointee became convinced that serious regulatory reform was essential. Carter's appointee, Alfred E. Kahn, who had been one of the leaders of the reform movement for years, was the first professional economist ever to serve on the Board. He was an eminent microeconomist and fully dedicated to reducing or eliminating much of the regulatory apparatus clinging to transportation and other industries such as power and communications. He had even written not long before his appointment what has become a classic book entitled aptly *The Economics of Regulation*.¹¹³ Kahn was the right man in the right place at the right time with the right support to persuade not only the Board and its staff,

a process already begun before his arrival, but as a witty, energetic spokesman for a cause with a solid analytical and empirical underpinning, he was able to convince many in Congress as well. He enjoyed wide popular support with the press and TV. The message they sent to the public also helped persuade Congress.

Much hard work was yet involved. But with strong bipartisan support from a Republican and a Democratic President and the leading Democratic Senator, legislation was prepared that even the industry, reluctantly to be sure, agreed upon or at least acquiesced to once convinced that some kind of reform was inevitable. In fact, substantial freedom of entry, encouragement of rate discounting and other breaks in the cartel-like arrangements of the previous Boards were instituted before legislation, some of which were of questionable legality. Under the broad terms of reference of the original Act, however, they seemed permissible although inconsistent with past precedents of the Board. Several carriers went to court to protest Kahn's initiatives. Passage of the *Airline Deregulation Act* in October 1978 ended these issues. The most sweeping deregulation bill to be passed up to the present, it was put in place, with several transition phases, one of which included the elimination of the regulatory agency itself, on January 1, 1985. This was the first and only "sunset" provision of any transportation legislation to date.

As far as intercity passenger transportation is concerned, the largest proportion of commercially offered service (some 90 percent of intercity passenger-miles) and the fastest growing segment had been relieved completely of economic regulation. This was accomplished with a speed and thoroughness that was totally unexpected even by those most ardently in its favour. Clearly this reflected not only a set of mutually favourable circumstances, but also recognition that deregulation, or at least regulatory reform, was an idea whose time had come. Just as the 1930s was the hey-day of the rise of regulatory commissions, so the late 1970s represented the time of their demise. This was at once a cheerful and sobering fact and one that haunts supporters at the present, some of whom fear a reverse trend toward some form of re-regulation.

BUS REGULATORY REFORM

The last of the transportation industries to yield to the reform movement was intercity bus transportation in 1982. The industry itself had none of the glamour of the airlines. Indeed in the U.S. it still has a kind of seedy

reputation typified by such stage and movie productions as "Bus Stop." Used primarily by the poor, indigent, aged, students and others unable to afford other means of travel and mainly where rail service is unavailable, it is the poor sister of intercity passenger service. The volume of its business did not change much between 1970 and 1990 for the nation as a whole, and the number of communities where buses once stopped appears to be dwindling rapidly from about 20,000 in 1970 to less than half this amount today.

Its share of intercity passenger-miles, including private automobiles, fell from 2.12 percent to 1.17 percent between 1970 and 1988. For the number of trips with round-trip distances greater than 200 miles, bus was the lowest in a modal split including auto, bus, rail and air in 1977 with less than 3 percent.¹¹⁴ By virtually any statistic, national bus transportation looms small. Even as a proportion of strictly commercially provided service, it is not large although almost double the rail (Amtrak) share in terms of intercity passenger-miles — 6 percent compared with 3.4 percent for rail.

The ICC regulation never seriously changed the structure of the industry from that which it inherited in 1935 when bus transport was included in the *Motor Carrier Act* without any differentiation from regulations pertaining to trucks. Bus regulatory reform was discussed from the late 1970s on in Congress and elsewhere, but it had none of the flavour and fervour of the debates, hearings and media hype that attended airline and truck reform. Since much of bus transport was intra-state, the previous federal regulations had more limited jurisdiction and impact. Thus the reform bill, the *Bus Regulatory Reform Act of 1982*, contained an amendment to the *Motor Carrier Act of 1980*, which freed up much of the trucking industry from economic regulation that specifically pertained to bus transport. This enjoined the states to cooperate on matters of bus transportation to accomplish the purposes of the Act, established procedures to ensure that this would happen and to prevent nullification by state regulatory actions.¹¹⁵ Both entry and exit requirements were eased, operating restrictions were removed that placed unnecessary burdens upon interstate commerce, and rate-making freedom, initially within a range and later removed entirely unless shown to be predatory or discriminatory, was granted. The ICC retained authority to make whatever distinctions were needed in defining "unnecessary burdens," "predatory," "discriminatory" etc.; to authorize new entry on fitness grounds only; and most importantly, to be able to overrule state authorities that "unduly" restricted discontinuance of purely intra-state operations.

The results of this Act were something less than dramatic. Greyhound retained its dominance and in fact gained market share for the nation as a whole for its total operations. All carriers took advantage of the easier discontinuance provisions so that the number of stops made, particularly at small towns, dropped sharply. Greyhound has suffered from severe labour problems since the early 1980s but began a system of franchising to reduce costs, increase its scope of operations and improve competitiveness against other bus operators and even other modes, mainly Amtrak.¹¹⁶ Service reductions throughout the industry do not seem to have been a result of reforms in 1982 since they had been going on for at least seven years prior to the Act. At most "the reform of motor bus regulations accelerated a previously established trend."¹¹⁷ Even so it is believed that "While intercity bus service has continued to decline nationwide, small communities receiving [Greyhound] franchise service have had relatively constant, or increasing, levels of service during the 1982 to 1986 period. . . . the early experience with franchising is . . . encouraging to those concerned about maintaining bus service to low income and elderly residents of small communities."¹¹⁸

Continued labour strife at Greyhound has however clouded the picture especially with respect to profitable private bus service to small communities and within rural areas, thus accentuating their isolation. Numerous local initiatives have been undertaken some with state and federal support and some privately financed by local charities as part of assistance to the indigent and elderly. Major intercity bus traffic has been affected as well but should continue to offer alternatives to Amtrak and, to a lesser extent, private cars and airlines. Any major resurgence of competitive strength however will await future conditions and is highly problematical.

Nevertheless, for the vast majority of small towns and most of the rural areas without air or rail passenger service, service by small or medium-size passenger vehicles at least to the nearest junction with an intercity carrier, probably also bus and Greyhound or one of its franchisees, will be the only alternative. In many cases it will have to be publicly financed.

Regulation of intercity bus service has never been its major problem. Nor can regulatory reform provide much of a solution. Economic forces will decide, and what reform has been implemented thus far will give these more latitude. To the extent that the needed travel by bus is neither for business nor vacation, it will take on the aspects of a social service

wherever alternatives are unavailable. Some form of non-user support will then be required, and many ingenious approaches have begun especially at the state and local levels. This implies some form of subsidy beyond the federal system. This also brings up the remaining large problem of direct subsidy to intercity passenger transportation, namely Amtrak.

With the *Bus Regulatory Reform Act* the period of legislative pro-competitive reform at the federal level ended in transportation. The most radical was in the airline industry but major significant changes were related to freight transport which indeed had precipitated the movement. Less complete deregulation occurred in truck and rail freight with the *Motor Carrier Act* and the *Staggers Rail Act*, both enacted in 1980. These two modes, omitting the specialized oil pipelines, accounted for almost 80 percent of intercity ton-miles in 1980. The inland waterways were never significantly subject to economic regulation in the first place nor were contract, exempt and private carriage by truck. Thus, for all intents and purposes, the entire freight industry was permitted to operate under the checks and balances of market forces with few remaining restraints. Certainly they were freer than ever to respond to economic and technological changes pretty much as they saw fit. Yet the ICC was retained as an entity to carry out certain reduced functions according to previous regulation but with new duties under reform. Other less significant industries were also reformed along these lines between 1977 and 1982 such as air cargo, air freight forwarding and household goods.

All commercial passenger service was substantially or completely deregulated by 1982 except for rail. The story of subsequent developments for bus has been noted above. The consequences of airline deregulation are analyzed in Section VI along with developments during the 1980s for intercity rail passenger service. What remains in the present section is a discussion of the development of the Amtrak experiment up to 1980.

THE CREATION OF AMTRAK: A DEVIATION FROM PRO-COMPETITIVE POLICY?

Rail passenger travel plummeted from 77 percent of all for-hire passenger service in 1929 to barely seven percent in 1970, the last year of rail owned and operated service. Worse was that its regular and large losses were jeopardizing profitable rail freight service. In one sense, relief for the railways from something they had been unable to make profitable and regulation

would not allow them to abandon, *increased* their ability to compete in the freight market. Greater freight rate reductions could take place without the need to support passenger service by cross subsidy, and the increased cash flow, without the out-of-pocket loss from the passenger business, would permit greater investment in more efficient freight-carrying capacity. The higher rate of return on assets devoted solely to commodity movement would also attract more capital. All these combined would create the opportunity for rails to compete more effectively with trucks, barges and pipelines over a broader range of goods traffic. In this sense the policy was pro-competitive and consistent with the growing concern with efficiency and market forces.

On the other hand, Amtrak represented a move in the direction of direct subsidy for a service in competition with privately operated, for-profit firms. To be sure, it could be argued that highway users were subsidized indirectly through failure to pay full user-charges for the publicly provided right-of-way. Airlines received a similar indirect subsidy for the airways and airports. These issues were harder to demonstrate certainly in terms of the amount of subsidy each competitor received, especially after the creation of the trust funds into which went some receipts referred to as "user-charges." Nor is it yet clear whether trucks and buses pay either their "fair" share or "efficient" share for their use of the multipurpose highways; similarly with airways. In any event, the subsidy to Amtrak was direct, obvious and politically charged. There was nothing ambiguous about it at all. In this sense, it was *anti-competitive* and non-market determined: hence the question mark in the title of this subsection.

The story of Amtrak and its formation has been well told.¹¹⁹ It will be briefly summarized here as part of the reform movement or at least one of the major consequences of reform that need to be examined especially in the context of passenger transportation.

As already argued, rail passenger service was taken from the railroads to preserve rail freight service, without much protest! However there were alternatives other than Amtrak. One widely recommended was to let inter-city rail passenger service die the natural economic death that seemed to be inevitable if market forces were to prevail, as Hosmer had argued. Another was to reimburse the railroads for the deficits from passenger service and continue as before allowing some abandonment in obviously hopeless

cases where “need,” not defined in economic terms, seemed to warrant. The government could simply nationalize the passenger component of rail operations or form a quasi-public corporation to do the same thing, but only after paring the service down to what might conceivably become a profitable or break-even service. What finally emerged was the latter — in part because surprisingly strong public protests influenced Congress not to let even the intercity service die. Part of this strong support for retaining something the public did not use very much was undoubtedly the experience of increasing numbers of people with European passenger service and the widespread publicity of the Japanese new technology and operations of the “bullet” trains. The possibilities of new high-speed technologies were also well publicized by projects examined by the government under the *High Speed Ground Transportation Act of 1965*. Thus, many felt it was premature to let rail passenger service disappear. They felt, on the basis of little evidence that the railroads had deliberately discouraged passenger service to be rid of it and that lurking out there somewhere was a technology that could revolutionize the service and make it economically viable. These views are still prevalent today. It was therefore the duty of the government to take the lead in developing such new techniques. If Japan, France and even England can do it, why cannot the U.S.? Some prestige was thus involved in Amtrak’s formation. At the same time more legitimate concerns were being expressed concerning environmental consequences of continued predominant reliance upon automobiles and airlines to provide nearly all the intercity travel.

Thus the *Rail Passenger Service Act of 1970* was passed, over the objections of the Council of Economic Advisers which had played some role in the development of the economic case for transport deregulation. There were other objections but President Nixon signed the bill without much apparent enthusiasm. He was reported to have been undecided whether or not to veto the bill up to the moment of signing.

The Act declared that rail passenger service was a “necessary part of a balanced transportation system” and therefore should be preserved. Furthermore it provided more freedom of choice and would help to alleviate highway and airport congestion. A separate corporation was created, the National Railroad Passenger Corporation, (later known as Amtrak) which explicitly was not an agency or establishment of the U.S. government but rather a private, for-profit corporation supported by the government.¹²⁰ The

Act required a basic system to be established in short order. Service beyond the basic system could be provided if deemed “prudent” or if a state, local or regional government body requested it and was willing to bear two thirds of the losses.

Amtrak contracted with the railroads to relieve them of passenger service in return for payments based upon each road’s 1969 passenger deficit. With equipment from the railroads as part of their payment, some cash from the government and the railroads too, as well as a basic plan that pleased no one, operations started under new management, so to speak, on May 1, 1971. Nearly half of the passenger trains in existence previously had been dropped, the “basic” system was national in scope — even connecting points where traffic potential was minimal — and the retention of many trains was due to political pressure. Operating with former railroad equipment 21 years old on average, relying on the railroads to actually run the trains under contracts which provided no incentives to give decent service and using the former railroad personnel unable to perform efficiently before, it is scarcely to be wondered that things did not go well. Table 4 summarizes the operating results through 1980.

Table 4
SELECTED AMTRAK OPERATING STATISTICS, 1971-1980

Year	Revenue passengers carried (millions)	Revenue passenger-miles (billions)	Deficit (\$ million)
1971 ^a	10.6	2.0	55
1972	16.6	3.0	148
1973	17.0	3.8	159
1974	18.3	4.3	273
1975	17.4	3.9	352
1976	18.2	4.1	441
1977	19.0	4.2	552
1978	19.2	4.2	582
1979	21.5	4.9	595
1980	20.8	4.5	599

Source: Annual reports of Amtrak, 1971-1983; also, Association of American Railroads, *Railroad Facts* (Washington, D.C.: the Association, 1983), p. 61.

a Reflects operations for the period from May 1 to December 31, 1971

To say the very least, the results are unimpressive. Comparison with the late 1960s and 1980 indicates the number of passengers to be barely one fourteenth, the number of passenger-miles, one third of those pertaining some 12 years earlier whereas the deficit is almost 50 percent higher. To be sure, the number of trains have been slashed but not nearly proportionate to the traffic reductions. The higher deficit, even considering inflationary levels, was still worrisome because revenues and costs should have risen no more than equally. Indeed with any improvement in efficiency and rational pricing, the gap between the two might well have fallen. Even with all the caveats, the service continued to be dismal during most of this period.

Yet there were several favourable signs. New equipment began to replace the antiquated materials foisted upon the Corporation in the first years. This equipment was less subject to breakdown, easier and cheaper to maintain and more fuel efficient. The first oil shock in late 1973 and the vigorous campaign to conserve gasoline and use public conveyances doubtless contributed to the ridership spurt in 1974 and again in the second oil crisis involving the Iranian revolution in 1978 which caused oil prices to more than double again. This added to the argument that the fuel efficiency of rail passenger transport was probably worth preserving in an environment of continually rising fuel costs, a proposition that was not disproved until the recessions of 1981-83. Finally, in 1976, the 4R Act gave the Northeast Corridor Boston-Washington main line and several branch lines to Amtrak. With its first ownership of right-of-way, it could provide its own service without contracting with the railroads and could improve the road bed to accommodate higher speed trains. Indeed the 4R Act mandated such improvements and ordered reduction in trip times as well. The Acts pertaining to Amtrak of 1978 through 1980 saw greater concern for cost control, elimination of under-used capacity and worry about the effects of subsidized Amtrak on bus operations. Goals were set to improve on-time performance, raise the average speed of the whole system, increase the ratio of revenues to operating costs at least to 55 percent by the end of fiscal year 1985 and in general to shape up.

In 1980 the newly elected Reagan administration was expectedly hostile to anything smacking of subsidy and openly attacked Amtrak's high visibility in this regard. After almost a decade of operations, Amtrak had gone from a situation where little was expected of it to one where nothing was expected of it. Yet the annual subsidies continued to mount. Hosmer was looking better all the time.

TWO STUDIES REGARDING THE U.S. TRANSPORTATION SYSTEM IN THE LATE 1970s

Concern over Amtrak led to the hope that perhaps the situation could improve if there were better coordination between rail and bus passenger transport. Coordination is a word used so often as a goal of policy that it ceases to be much more than a pious hope, especially in a system that has from the outset insisted upon separation of modal ownership. Nevertheless, the DOT was asked to report on just such a possibility. The report, entitled *Report on the Potential for Integrating Rail Service Provided by the National Railroad Passenger Corporation with Other Modes*, appeared in 1976 and answered with a resounding "No."

The reasons given for such a negative conclusion reside in the barriers to cooperative integrated service, the most important of which is the inherent antagonism between the two modes. Most bus operators, especially those providing parallel service with Amtrak over at least some of their routes, view Amtrak with suspicion and as an unworthy subsidized competitor. There was little desire, at least at the time of the Report, for closer relations between the two modes.

There are of course inherent reasons why intermodal public transportation "involves an added degree of passenger inconvenience when compared to single mode travel."¹²¹ The added inconvenience arises because the intermodal traveller is usually required to "(1) change stations; (2) endure lengthy layovers as a result of poor connecting schedules; (3) purchase separate sets of tickets; (4) transfer baggage; and (5) seek schedule information from multiple sources."¹²² The Report examined each of these in the context of Amtrak-bus relations, institutions and attitudes. Without going into any details, suffice it to say that the Report found the sum total of these problems virtually insurmountable and the expected traffic potential so limited even if they were overcome, that the effort and added costs were scarcely worthwhile. This conclusion is relatively easy to accept even though there are clearly specific situations where such coordination is possible and economically feasible. There appear not, however, to be very many of them.

The second report is far more substantive and comprehensive. This study, entitled *National Transportation Policies Through the Year 2000*, was commissioned by Congress in 1976 and directed to conduct "a full and complete investigation and study of the transportation needs and of the resources,

requirements, and policies of the United States to meet such expected needs.”¹²³ To implement the mandate, a National Transportation Study Commission was established, consisting of 18 members (and nine “former members”) of which eight were businessmen and the remainder Congressmen or former Congressmen. A large staff was assembled whose findings were subject to approval by the Commission as were the details of the final Report. The chairman noted that the recommendations were approved unanimously by the 18 members.

The timing of the Report was unfortunate since it overlapped the period of most rapid introduction of new legislation, 1976-1979, when it appeared. It was therefore unable to influence the legislation nor did it have the advantage of knowing what kind of legislation would emerge with which it could agree or disagree, except for the airlines. It is lengthy, loaded with data and analysis based upon considerable sponsored research.

Its analysis of transport “needs” required future projections of the state of the economy starting with base year data for 1975 and ending in 2000. Much activity was devoted to estimating the economy and the associated transport levels at that time. From this were deduced the needs which more or less became confined to capital requirements and the policy changes in some sense “necessary” to accommodate the estimated needs. Since the relationships between the state of the future economy, the number of associated ton- and passenger-miles and the capital plus policy changes essential to ensure achievement of the projected transport output are not amenable to quantification without some pretty heroic assumptions, the Report reverted to much descriptive material. When it came to numerical estimation, a lot was left to be desired regarding methodology, data and the (usually) implicit relationship between the end results and the needed transportation resource inputs (capital and energy). The subjectivity reached even higher levels in relation to the linkages between the (over 80!) specific policy changes from the so-called *status quo* situation (itself then undergoing rapid change). Somehow each recommendation was presumably necessary to produce the needed transportation services of all kinds. Such recommendations referred to government organization, economic regulation, non-economic regulation, ownership patterns, financing, pricing, taxation, planning and information, the fuel situation (prior to the second oil crisis in 1979) and so on.

While the scenarios were reasonable enough (mainly the low-growth scenario as far as real GNP, population and labour force were concerned through 1985), the use of the material and its enormous scope made much of the discussion either overly descriptive (though not uninteresting), facile, inconsistent or tedious. The research papers contain better and more useful analysis. The attempt to put them all together in a single, even lengthy volume does not come off.

For what it is worth, the policy recommendations come down squarely in the middle of what was then occurring in the economic policy realm — namely, more reliance upon market forces and free enterprise but with a new super commission to include the ICC, the CAB (soon to disappear) and the Federal Maritime Commission (FMC). Few could disagree with the conclusion that “Over the long term, complete revision of existing Federal economic regulation of transportation should be accomplished. Revisions should be undertaken following careful but expeditious economic analysis of their consequences along with assurances that protect the public from possible adverse effects implicit in such revisions. Basically, such revisions should encourage efficiencies in operation and energy consumption and provide wider latitude in entry, exit, route choice, price, and quality of service by permitting firms to make their own decisions subject to general laws and rules designed to protect the public interest, such as energy conservation legislation, antitrust laws and insurance coverage. Transportation firms should be permitted to merge, however, subject to the same antitrust policy concepts as other sectors of industry, as a minimum, once regulatory reform is implemented. In applying antitrust policies, consideration is necessary for the interconnected, international, and interrelated nature of the transportation industry.”¹²⁴

The Report was something of a disappointment but it did reflect the mood of being between two policy regimes. It had the main thrust right but worried too much about how to get from where we were in 1976 to where we, in some sense, were projected to be, equated here with where we *ought* to be, by the year 2000.

6. A DECADE OF EXPERIENCE UNDER REFORM

The last year of major legislative change was 1982 when bus transportation was granted more economic freedom from federal regulations. No new reforms along these lines have been made since possibly because, aside

from Amtrak, there was not much left to do in transportation regulation at the federal level. To be sure, the ICC was left with some new duties and reinterpretation of old ones but in intercity passenger transportation it had few remaining duties or authority of much consequence. The CAB was gone by 1985 and had few functions left except for winding down. Thus, aside from water carriage, ferry service and other aspects of state and local intercity travel, economic regulation had been "reformed" along pro-competitive lines.

There are two main questions to pursue in this section as we bring the story up to date. One is, what have been the consequences of regulatory reform for air traffic? We have already noted the situation with respect to buses. And Amtrak was not reformed, it was created, but we will examine what has happened to it as well. The second question is, depending upon the answers to the first, what should the federal government do next with respect to the transportation industries? We have some clues from the DOT's statement of policy in 1990, but the issues remain of what more or less needs to be done in the interests of efficiency including problems of externalities involving safety and the environment, as well as user-charges.

THE AIRLINES AFTER DEREGULATION

Even before the *Airline Deregulation Act* was passed late in 1978, the experiment in decontrol had begun in earnest under Kahn and to a lesser extent his predecessor. Fare discounts were not only allowed without the usual regulatory fuss but actively encouraged. Discount fever swept the industry. Over half the total fares were subject to discount by the summer of 1978. As a result, load factors jumped on average almost 10 percent from 55.5 to 60.7 percent for the year ending September 1978. Carrier profits nearly doubled over the same period. The demand elasticity had been effectively probed under competitive pricing and it had worked. The formal deregulation Act was passed in October of 1978 partly because of such effects although the votes were there before these results were in.

This auspicious beginning was not to last. The next year witnessed the second oil crisis with fuel costs doubling. Price competition intensified and new entrants appeared in many city-pair markets. The general fare level rose but not due to deregulation — or so several early studies concluded. In fact, the fare discounts and out-and-out price competition maintained traffic

quite well. The recessions of the early 1980s, associated with attempts to resolve the supply-side oil shock by stringent monetary policy alone, reduced air traffic demand and led to several bankruptcies of note (Braniff and Continental) and some of the new entrants did not survive. Overall losses occurred in 1980, 1981 and 1982, but as the economy revived in 1983 and began its long period of growth that ended in mid-1990, the domestic airline industry performed quite well. Competition in most markets intensified.

On a national basis there was a return to relatively high levels of concentration as mergers and bankruptcies continued, but market concentration route by route decreased on average in markets of all sizes and dimensions.¹²⁵ Recent results (July 1991) show 6 of the 12 largest carriers in bankruptcy. Four of these have filed since December 1990 and one, Eastern, has since ceased operations. Concentration at the aggregative level has potentially increased over the last year by a substantial amount depending upon how the bankruptcies are resolved. Except for Eastern, the current situation has not changed much since the other carriers are still operating and two of them, Pan Am and TWA, have mostly been in international business, which remains highly cartelized. The announced recovery from the recession that began in the summer of 1990 may ease the situation if more vigorous growth occurs than anticipated at present writing (July 1991).

However, most of the evidence up to the present shows that competitive forces are working generally as expected in the airline industry; namely, fares are more volatile in response to changes in supply and demand conditions; there have been more bankruptcies which is the normal process by which excess capacity or managerial ineptitude become resolved, reduced or eliminated in competitive markets; and a closer coincidence between rates and costs in specific situations has been achieved, although the evidence here is mixed.

At the same time there are certain other problems beyond those associated with recent bankruptcies. Price discrimination has apparently increased. Service quality overall has decreased as manifest by the increase in airport congestion which causes takeoff and landing delays. There is an increased sense of crowding aboard most flights as a result of fare discounting and the reduction in excess capacity through higher load factors which have been sustained above the 60 percent level in the late 1980s.

Some of these results are beneficial to most passengers. The enthusiastic response to rate reductions — even discriminatory ones — has helped fill up aircraft to a greater extent than before. For some, the crowding effect is unpleasant and they would be willing to pay more to have more space. To a limited extent this option is available with different classes of service (e.g. business, first class, etc.) but not always. It is doubtful that this inconvenience constitutes enough disutility to suggest that there are no overall net benefits. Kahn gives some evidence to this effect.¹²⁶ The increase in load factors is however more efficient and permits lower fares as was intended. The overall improvement in resource allocation is to be commended even if not a totally unmixed blessing.

The increased delays occasioned by the enormous traffic growth since deregulation, itself a function of competitive pricing due to contestability that produced lower fares on average and sharp discounts at specific times in particular markets, is related to the federal government's failure to encourage more airport construction at the state and local level or become more involved itself. Certainly there is no excuse for having supported a policy change that was expected to stimulate the use of excess capacity and to encourage the expansion and use of additional capacity in aircraft, and then not to have accommodated this with commensurate expansion of airway facilities which, from the beginning, have been a federal government responsibility. The excuse of budget deficits is a sham since such outlays are investments. The fact that no new major airport has been constructed since Dallas-Fort Worth in 1974, well over 15 years ago, suggests the source of congestion delay. This may now be on the way to correction but it is very late in the day. Rather than reflecting adversely on the wisdom of deregulation, increased traffic (the cause of congestion) reflects favourably on deregulation. The lack of facilities reflects mostly on the lack of vision of government transport investment policies.

The intensification of price discrimination is more worrisome. In part it reflects the continued concentration in some submarkets and the new dominance of certain major hub airports for the increasingly popular hub-and-spoke system by one or two major carriers. This is partly related to the failure of the federal government to enforce the antitrust laws thereby permitting a degree of concentration that would otherwise not have occurred. This failure was a deliberate part of the Reagan administration's policy of virtually ignoring antitrust enforcement during most of the 1980s as an

aspect of extreme *laissez-faire* and the belief that "government was the problem." As Kahn puts it, "the government clearly has neglected responsibilities of which it was never the intention of deregulation to relieve it."¹²⁷ This also applies to airway and airport capacity and enforcement of safety practices, including the failure to replace all the air traffic controllers fired in the early 1980s despite the large traffic and congestion growth.

The issue of discrimination is also, and perhaps mainly, due to the ability, through control of sophisticated computer reservation systems, to take advantage of varying degrees of monopoly power on specific routes to engage in "yield management" — which is a euphemism for trying to maximize net profits from each seat-mile on each flight. With profit limitation and/or *some* greater degree of competitive pressure, differential mark-ups over marginal cost are both essential and efficient, as Ramsey pricing theory indicates. It would be folly to deny it to the airlines when the *Staggers Rail Act*, which "reformed" the rail freight industry, virtually mandates such pricing in the interests of revenue adequacy. Yet the extent of such deviations from costs and the failure of the fare structures to pay attention to distance bothers many people. Vastly different fares for the same flight and service reinforce concerns about "fairness" or apparent equity. The original *Act to Regulate Commerce* arose out of just this kind of emotional response to long- and short-haul railroad pricing. At some point there is a trade-off between equity and efficiency, especially where the degree of fare differentials becomes so wide that it induces a customer backlash that could lead to some re-regulation however much economists may justify Ramsey pricing. It simply does not seem fair. The encouragement of more open access to computer reservations systems, more vigorous antitrust activity and expanded capacity of the infrastructure to attract competitive challenges to hub-dominance, might alleviate the extent of discontent and prevent more drastic re-regulation.

All things considered, airline deregulation has been at least a qualified success. The industry is certainly more competitive and efficient than it was. The benefits of improved efficiency have been distributed widely if unequally. There is little evidence of monopoly profits. Most other industries with these features would be let alone or tolerated for the degree of market concentration, industry practices and results that have thus far typified the airlines. Furthermore the prospects of even more competition from foreign airlines are promising. An "open skies" policy in Europe, scheduled for 1993,

could be matched by a similar policy in North America. The U.S. Secretary of Transportation is reported to have said that "America and Europe will one day negotiate a deal to allow their carriers completely free access to each other's markets."¹²⁸ It is therefore much too early to talk about reversing the past 13 years of substantial economic freedom in the U.S. airline market despite some real and potential problems none of which require major deviations from the promising trends already established. The federal government however must change its stance toward airport and airway capacity and efficient pricing thereof and more rigorously enforce antitrust laws. It should also support open-skies policies. Eternal vigilance is the price of maintaining competition. It is also worth it.

AMTRAK IN THE 1980s

One might wish that the intercity rail passenger transportation problem had been as well resolved as that of the airlines. Unfortunately this is not true although there are signs of improvement.

Since 1980 the number of passengers has risen less than 10 percent and passenger-miles about one third. The federal subsidy is estimated for 1992 to be about what it was 12 years ago but one third less than its peak of almost \$900 million in 1981. In short, it has shown traffic growth far below airlines, but above intercity buses and little reduction in the amounts required from the government. It still plays a minuscule role as a *national* carrier but is much more significant in the Northeast corridor where it carries about one third of the combined rail and air passenger traffic. Even here the proportion falls sharply to less than two percent if bus and automobile intercity traffic is included. After 20 years one cannot say that Amtrak has not lived up to the limited ambitions as envisaged by its supporters.¹²⁹ But it has done little more than hold its own if we discount, as is only fair, the initial five years as experimenting with inadequate equipment, despondent personnel and more than the usual start-up problems. It is clearly not a national carrier as over half of its business is done along the Northeast corridor.

As a short-term measure to relieve the railroads of their losing passenger service and at least salvage the freight business, it has been successful. As an experiment to see whether *traditional* rail intercity passenger service has a future on a national scale it has failed. It is clearly of some consequence in the Northeast and perhaps has a chance in several other corridors but that

seems unlikely. The existing technology has few prospects as either a major player along the limited number of high-density corridors or, more certainly, as a national entity serving hundreds of communities and cities, as is being attempted now. Indeed it is probably obsolete even as an important aspect of any specific high-density corridors. For these, there are the options of the higher technology that has long been in use in Japan and France and new technologies being developed such as magnetic levitation (maglev) and others.

High-speed trains "are an entirely new mode of transportation".¹³⁰ They run on exclusive rights-of-way, using different power systems and can achieve speeds of up to 300 miles per hour (mph). Amtrak's top speed is only 125 mph between New York and Washington, D.C., and this only after a \$2.3 billion investment in track upgradings, removal of all grade crossings, introduction of electric locomotives and improved signal and communication systems.¹³¹ Maglev techniques even by-pass the historic "flanged steel wheels on steel rails" which virtually defined the railroad concept. Not only is this a new technology but it is one that is changing rapidly. So far as one can see, Amtrak is unlikely to be involved with it in any meaningful way. In fact separate states have been examining these alternatives for up to ten years largely independently of Amtrak. They include Texas, California, Pennsylvania, as well as consortia of states such as the Midwest High Speed Rail Compact, the Coalition of Northeastern Governors. The apparent success of the TGV in France and the Shinkansen in Japan and the prospects of the recently opened German Inter-City Express on May 29, 1991 with maximum speeds above 175 mph and potentials above 200 mph have stimulated renewed interest in North America, although it may have much more limited application than in Japan and Europe.

There are several reasons for this. In the first place, distances between major urban centres, or areas where high-density corridors are possible, are much greater in North America than in Europe and Japan. If the competitive range for high-speed ground transport is taken as between 150 and 500 miles and requires a ridership of about six million passengers per year to break even, given the most likely combination of fare and cost level, there is only one corridor in the U.S. now generating that number of passengers by air (Los Angeles-San Francisco) and only four expected by the year 2010. Any high-speed ground transportation system would have to capture all of the existing air traffic and/or a substantial portion of air plus automobile and bus traffic

to be economically viable. Given the likelihood that speed, capacity, comfort and efficiency improvements would be occurring in the competitive modes as well through the year 2010, the chances of any large-scale, unsubsidized, rail intercity high-speed operations seem slender indeed.

Secondly, the European and Japanese systems originated in heavily travelled rail corridors that were approaching capacity and could therefore build incrementally upon a system that had already achieved relatively high speeds.

Thirdly, the systems in Europe and Japan were far more fully integrated into a national system of conventional rail and terminal networks for passengers which no longer exists in the U.S. to anywhere near the same extent.

Finally, the relative dependence upon rail for intercity travel was far greater in Europe and Japan than in the U.S. when the former began investing heavily in high-speed rail facilities. In fact, in the U.S. rail intercity passenger traffic was dropping rapidly and with it any major concern for its maintenance given the comparative advantage of automobiles and airplanes in the U.S. with relatively low fuel prices and ubiquitous highways and airways priced at less than full costs to users.

For all these reasons, the technologies for high-speed rail passenger traffic were developed and introduced far earlier elsewhere, Japan in 1964 and France in 1981 for the actual beginning of revenue service. Any transference of them to the U.S. will have to face up to substantially different circumstances here and to the historic policy of modal independence. There is simply no national constituency for implementing an almost wholly new transportation system that would compete effectively without a heavy infusion or commitment of federal resources at least to supplement state, local and private finance, and far more support than implicit in *Moving America* (see next subsection).

Amtrak seems remote from such considerations and federal involvement is not great yet since, except in the Northeast, the systems thus far envisaged are almost entirely intra-state. They are also enormously expensive. However, the TGV reports, up through 1985, internal rates of return for the Southeast system of "15% for the French Railroads and 30% for the community as a whole."¹³² A recent study by U.S. DOT relating specifically to maglev technology concludes that "it will be economically feasible to

construct a limited number of commercial maglev systems in the United States, starting in this decade."¹³³ This is only for a limited number of route-miles and, except for the costs involved (about \$50 billion), remind one of the early turnpike roads. Most reports are less optimistic about economic feasibility.

This is all highly speculative. As it pertains to Amtrak, however, it suggests that either its operations should be confined to the Northeast corridor and the track there upgraded to achieve significantly higher speeds using newer already proven technology (maglev is of course premature) or the experiment be wound down. If the only reason for keeping it going is that the value of the assets at liquidation would be far less than the obligations of the Corporation and thus the government would become a payer of last resort at an expense greater than the current subsidy, as implied by Tobey,¹³⁴ then the time has surely come to end it as gracefully, quickly and expeditiously as possible before even more capital commitments are made.

There may however be efficiency reasons for continuing at least portions of the Amtrak experiment. A recent route-specific analysis of Amtrak's operations, based upon a disaggregate intercity passenger demand model for 1977 calculated a measure of net social benefits (NSB). It concluded that "government subsidy may be economically justified on dense short-haul routes like those found in the northeast corridor, but subsidizing long-haul and low-density routes cannot be justified using economic efficiency arguments."¹³⁵ In fact the net social benefits are positive for the whole operation but only because of the huge net benefits from the Northeast corridor. One can, of course, quarrel with the demand model employed, the data available and the method used to calculate NSB. However the general findings seem reasonable enough and the whole approach, even with the acknowledged qualifications, verify earlier conclusions that the "benefit offered by rail travel is greatest on short routes between large cities where the primary beneficiaries are business travellers."¹³⁶

Though the system may not be financially viable, the efficiency criterion requires only positive NSB, although some of the social benefits could be captured on selected routes through some variant of Ramsey pricing to reduce the specific financial subsidy. This combined with continued improvements in cost effectiveness and productivity could turn the Northeast corridor into a financially break-even operation. If Amtrak is viewed as a basic

national rail passenger system, then financial viability and net social benefits are reduced. Few network economies or benefits exist in the sense that most of the routes outside of the Northeast corridor do not feed much traffic into it nor provide contributions to joint or common costs.

The experiment attempting to create a *national* system has not succeeded and shows little prospect of much improvement. As with the railroads and passenger service, a decent effort was made but to no avail. It should be stressed that ending Amtrak does not preclude development of high-speed rail in the U.S. along specific corridors. Indeed, that appears likely within the next decade. It will not however bear much relationship to Amtrak as originally conceived and thus far executed except for the Northeast corridor.

THE LATEST NATIONAL POLICY REPORT

In February 1990, the Secretary of Transportation released the most recent statement of transportation policy under the title *Moving America, New Directions, New Opportunities*.¹³⁷ This is based on several assumptions. First that the basic U.S. transportation infrastructure is complete in the sense that massive extensions of it are no longer necessary nor, with few exceptions, would they prove to be economically feasible. The system is mature. It needs lots of work however because of inadequate maintenance and increasing capacity constraints. It is now more important not only to maintain the system but to use it more efficiently: Indeed economic and efficiency criteria are everywhere in evidence throughout this Report. Thus, the way to use the infrastructure more efficiently is through sensible pricing. This covers a host of possibilities including peak/off-peak pricing, cost-based pricing including full user-charges for use of publicly provided facilities including externalities, elimination of subsidy, modal coordination, modal equity of treatment among modes and so on. New technologies are emphasized including high-speed rail, intelligent vehicle/highway systems, research to create and disseminate new technologies and demonstration projects for new techniques. To improve efficiency, intermodal equity and allocation within transportation, the remnants of economic regulation of rail and truck freight transportation are to be removed. As far as this Report is concerned, the regulatory reform movement has not only been successful but is continuing. Even freer markets and more of them are sought in transportation.

As would be expected, safety and defence needs are emphasized, but couched in terms that are not inconsistent with efficiency criteria, namely that, given the goals, they are to be achieved by least-cost techniques. Even the size of the estimated defence needs will be scrutinized to minimize negative economic impacts.

Another main assumption underlying the Report is that the federal role should be limited in various ways. Federal funding will be available only to projects of truly "national significance."¹³⁸ Furthermore, "the increased responsibilities and capabilities of the State and local and private sector partners" must be recognized.¹³⁹ This applies not only to maintenance but to other projects as well. Greater use must be made of user-charges and innovative financing techniques by state and local governments and the private sector. "There is widespread recognition that the United States will need to rely more on the private sector to finance facilities and services in virtually every area of transportation . . . [G]reater use of innovative financing approaches, including private financing initiatives and joint ventures by private companies and State and local governments"¹⁴⁰ should be encouraged. "State and local governments have already assumed greater responsibility in transportation. This can and should continue."¹⁴¹

Along with reduced largesse, the federal government will relax constraints upon how the reduced funding may be used by the state and local authorities. "It is Federal transportation policy to . . . move toward greater flexibility in use of transportation funds at all levels of government for facilities that enhance access and improve connections."¹⁴²

The limitation on the federal role in finance will however be partially offset by activities that enable state and local governments as well as the private sector to plan and allocate resources, wherever or however derived, more effectively and especially where they fill gaps in the national system.¹⁴³ Federal outlays will also be increased to assist "research and technology projects" in a wide variety of ways such as providing seed money, maintaining a knowledge base, promoting "adoption of new technologies"¹⁴⁴ and so on.

In short, beyond the reduction of federal activities already implicit in the deregulation that has occurred since 1976, further reductions will take place in the funding of new transportation assets even though future needs will

be substantial. The federal role will be more one of monitoring resource use, assisting in various ways to enable others to tap financial resources and use them effectively, promoting innovations, education and competition in transportation and, in general, playing a mostly hands-off, chairman-of-the-board type of function. This contrasts with the previous role since the Great Depression when vast amounts were paid for highways, airways and other aspects of the infrastructure to say nothing of the outlays for the several regulatory commissions. In a sense, the federal government is putting the states, local governments and the private sector on notice that the big job of infrastructure building is done. It is up to them to maintain and manage it properly.

The federal government will assist with gap closing, maybe some infrastructure extension but mostly with supervision and some managerial help. This has been referred to as "a vision of federal policy that is narrowing, retrenching, retiring."¹⁴⁵ It is also probably true that this withdrawal is a recognition of the enormous costs in money, accidents, environmental and aesthetic degradation occasioned by much of the economic regulation and infrastructure building of the past century and a half or longer. One suspects that it is mainly motivated by the federal budget situation. Even so, it is a reasonable retreat. Market forces are doubtless better users of the facilities than regulation and there is little doubt that much excess capacity exists throughout the system's right-of-way, despite serious congestion in some areas. Proper pricing and some efficiently conceived new capacity will probably suffice for the next decade or so especially if bolstered by new technologies that are not outrageously costly. We have come full circle from the days when the federal government was reluctant to spend anything on transport right-of-way except for projects of "truly national importance" and left most of the regulation and investments in transportation to the states and local communities, through the activist period when the opposite applied, to 1990 and back to *Moving America*.

THE UNFINISHED BUSINESS OF REGULATORY REFORM

The unfinished business of economic reform of intercity passenger transportation involves a package of at least three major ingredients: first, a more conscientious effort to price publicly provided infrastructure more effectively and consistently across modes so that competition can play its role in the passenger business. Second, from the very beginning, there should

have been a moratorium on mergers, except for the most compelling cases of, for example, inevitable bankruptcy or termination of an active competitor in order to maintain as many traveller options between city-pairs as possible and, hence, give competition and market forces a real chance to see how effectively they can work. Finally the regulatory authority needed to be stripped of all powers to prevent freedom of exit and entry and to price in any way the firms saw fit so long as no antitrust laws were violated. In essence, treat the ICC just as the CAB was treated.

This was not done in intercity surface passenger transportation. In a real sense the experiment in regulatory reform was not allowed to work out with respect to passenger traffic as it was with respect to freight. Perhaps this relates to the fundamental differences between the two markets noted earlier: moving people requires a higher degree of manifest public concern for safety, comfort and convenience than does moving things. Market forces are not to be trusted to the same extent where people are involved. The failure to allow market forces to determine the quality, quantity and price of intercity rail transport reflects this approach which surprisingly did not emerge with respect to air transport. Thus, other factors must be at work.

BEYOND REGULATORY REFORM

Among these other factors are the new priorities which began to emerge in transportation policy even before the first deregulation bills were passed. Protection of the environment, transportation safety, energy use and urban sprawl were issues that preceded regulatory reform and remain even more urgent today, especially as the U.S. Congress debates what has been called the "first major overhaul of the federal transportation system in 35 years."¹⁴⁶ Part of the rationale for Amtrak was that rail intercity passenger travel was more fuel efficient, safer and less polluting than automobiles or airplanes. Thus there would be substantial external benefits to a more extensive use of railways which would more than justify the meagre subsidy. It was also believed that autos and planes were heavily subsidized in many different ways most evident through inadequate user-charges and uneconomically low fuel prices which encouraged excessive use of scarce fossil fuels, created congestion, pollution and suburban sprawl and caused more deaths and injuries. Thus the highway and airway ages were seen by some as artificial creations of a public policy that has warped the outcome

against alternatives such as railroads and buses, both of which could move the same number of passenger-miles at incredibly lower real "costs" in all of these dimensions.

There is, in all this, no recognition of the value to the user of the enormous quality differences among the modes which is especially important in passenger transportation compared with freight. A ton-mile of cargo moved by rail or truck may not differ much from some shippers' point of view but a passenger-mile by bus, Amtrak or TWA has a huge difference in "value" as average passenger fare differentials suggest. Demand considerations cannot validly be neglected in efficiency determination as the centrally planned economies learned the hard way. Again, this type of thinking about efficiency only in relative cost terms is partly a product of the failure to distinguish between freight and passenger transportation.

The environmental concerns however highlight the need for a more careful examination of user-charges, including amounts necessary to compensate for congestion and pollution. This is no easy task. Yet the importance of efforts to do the best possible even with the existing inadequate information cannot be over-stressed. In addition, it would be worth considerable investment of resources and efforts to obtain information to enable more reasonable estimates of the full costs, including the externalities, of private and commercial use of the highways and airways, by vehicle and plane type, and assess these accordingly. It is doubtful that this would significantly change the modal composition of travel demand but it would provide a substantial incentive to automobile, truck and aircraft producers to enhance fuel efficiency, develop alternative fuels and reduce pollutants. This would also improve the performance of markets in achieving a true least-cost outcome in the economic sense. In fact this was what the regulatory reform movement was all about in the first place. It required full user-charges and maintenance of traveller modal and intermodal options for the effective functioning of commercial transportation markets freed from economic regulation. This was seldom made explicit. As a consequence, maximum efficiency has not yet been realized in the case of intercity passengers.

The most recent concern is about growing congestion of certain major areas of the highway and airway systems. However, these apply mainly to selected airports and to highways in and around major cities. There is plenty of excess capacity elsewhere in the intercity rights-of-way beyond

the urban confines. Yet there are plenty of reasons for concern for intercity transportation as well. Most intercity travel originates or terminates at an urban centre or airport hub where congestion is likely to become increasingly severe. Up to now, the approach has been to build more physical capacity. But there is now more opposition to this for safety reasons, because of excessive costs in already built-up areas, pollution concerns including noise, visual pollution as well as noxious emissions and others. No new major airport has been completed since 1974 in part because of strong community opposition — the NIMBY (not-in-my-backyard) effect. Paving more urban freeways or by-passes runs into similar opposition and when they do get built usually only arrest congestion for a short time. The effort is now focussed on how to get more transportation out of the roads we have already built. “We have poured enough concrete,” as Senator Moynihan vividly proclaimed in the context of the pending highway bill.¹⁴⁷ Indeed, using the existing infrastructure more efficiently is one of the goals of regulatory reform. It also requires full user-charges, time-of-day pricing and other schemes already noted. These also apply to the *urban* air and highway networks along with many other necessary measures to ameliorate the so-called urban problems.

Thus the emphasis upon improved maintenance and more efficient use of the existing infrastructure makes special sense at the current juncture. It will however take much more than simply eliminating the kind of economic regulation that involved the ICC and CAB and placing blind reliance upon even competitive market forces, important as these may be. The role of the federal government in intercity passenger transportation is far from over. Indeed, it may increase sharply if high-speed ground transportation is to proceed much further. Ironically the role will shift from deregulation of air and bus service to promotion of some “rail” (higher speed, of course) intercity travel, the shrinking and confining of Amtrak to the Northeast corridor and the more active promotion of competition within the air and bus industries.

EFFICIENT INFRASTRUCTURE PRICING: HIGHWAY USER-CHARGES

More than deregulation and greater reliance upon competitive market forces are needed to improve transport efficiency and performance. It is the duty of government to ensure that most if not all transport markets are and remain competitive, contestable and, in the economic sense, “free.” There is little point in substituting a private for a public monopoly if one cares

about "the public interest." Constant efforts are required from these agencies of government concerned with antitrust or anti-combines policies to ensure free, open and contestable markets in the transport sector following regulatory reform. But even free markets are not enough. Since much of transport infrastructure is publicly provided, no automatic, free market mechanism will ensure that it will be efficiently priced to its users, the providers of transport services whether public or private. Thus efficient infrastructure pricing is equivalent in importance to deregulation and maintenance of competition if we seek efficiency of production and consumption in transportation.

The favourable economic outcomes resulting from a largely pro-competitive market, require that suppliers of passenger services pay the full costs borne by society in producing such outputs. The marginal social costs of production is the "price" that needs to be charged for all goods and services, including those of publicly provided infrastructure.

In all areas of production, there are problems in measuring precisely what constitutes marginal social or even marginal private costs. Where externalities are involved, the difficulties of measurement are substantially increased. In transportation the situation is further complicated because much of the infrastructure is publicly provided, maintained or improved from its natural state as in the case of waterways and ports. The "services" of such infrastructure are not only hard to define, but difficult to cost and hence to price properly, whether used for commercial or own-account purposes. The fact that some modes (e.g. rail and pipelines) own their own right-of-way and must somehow price it correctly in the product sold, does not solve the problem; it merely changes the locus of responsibility for doing it.

Within the entire transportation industry, including passenger and freight commercial and non-commercial, a set of at least reasonably plausible cost-occasioned user-charges is indispensable. Yet few attempts have been made to develop such estimates or even to build, maintain and finance highways using efficiency criteria as at least part of the input. It is thus difficult to establish whether particular highway user-groups pay either their fair or an efficient share of the total costs of the highway system, although various partial studies in both Canada and the U.S. tentatively have concluded that heavy trucks seem to pay less in user-charges than the costs they occasion, while passenger cars and buses pay more. But without improved estimates, data and methods even this tepid statement is largely guesswork.

More attention has not been paid to this important aspect of efficiency for many reasons. These include socio-political rationales that stress the public nature of highways and the need to have them "freely" accessible to all, that many of the benefits accrue in the form of greater national unity, national defence, generalized stimuli to regional and national growth and other such quasi-public goods. Hence no specific user-charges should be made since the benefits are widely diffused. However, highway investment, use and maintenance utilize resources that could have been employed to produce other goods and services of equal or greater value. To prevent over-investment and excessive use of highways, or anything else for that matter, requires the imposition of efficiency criteria. When competing wants are considered, so long as we live in a world of relative scarcity, resources and their use must be priced appropriately if they are to be provided in the relative quantities demanded at the least social cost. Thus highways produced in amounts decreed by non-economic criteria and financed from general taxation are likely to absorb excessive amounts of real resources. Some linkage between infrastructure production, financing and user as well as non-user costs and benefits is clearly required.

Another reason why these linkages have not been instituted in Canada and why there is no formal relationship between road taxes and expenditures is because data have only recently become available in Canada.¹⁴⁸ More important is that there is no general agreement on which of the various methodologies should be used to allocate highway costs even if all the appropriate data were available.¹⁴⁹ Considerations of efficiency easily become intermingled with equity, subjective judgement, costs occasioned versus benefits received, appropriate units of measurement and so on.

In short, there remain many controversial issues in this area before, as Nix puts it, any "cost-allocation study could be undertaken for policy guidelines."¹⁵⁰ This is not a counsel of despair, for various measures can be implemented that move in the direction of greater efficiency. The potential gains from so doing are substantial. The current financial administration of highway systems in Canada and the U.S. so disguises the real resource costs attributable to different users and non-users that efficient production and use of highways and substitute modes are seriously distorted. Since highways represent a significant portion of each nation's capital stock, and annual investments and users thereof represent a substantial share of national income and expenditures, it follows that there is probably considerable economic waste from inadequate user-charges.

ESTABLISHING LINKAGES

Linkages between highway expenditure and road user-taxes or charges must be made. A more efficient set of user-charges for highway systems and an accompanying scheme of fuel and other taxes and fees must be designed to recover the marginal social costs of highway use from each user or user classification.

To begin with it is necessary to define a road user-charge with some care. Sales taxes on vehicles, trailers, parts and other transport equipment and income taxes on commercial providers of transportation cannot be construed as highway user-charges because they are unrelated to use of the highway. Even fuel taxes cannot logically be viewed as charges for the use of the highways *unless* they are established differentially higher than any existing general sales taxes and the additional proceeds specifically linked to highway costs occasioned by the users.

For existing highways, the costs to be recovered from users are the annual maintenance, repairs, superintendence, policing and administration cost that vary with highway use. Past capital costs need not be recovered — they are sunk costs. New highway investments should be made on the basis of well-known benefit-cost criteria.

(a) *Highway user-charges*: Even without complete information, user-charges can be established that will allow users to calculate the costs of additional vehicle trips that will specifically include the additional wear and tear, etc. on the highway itself. If congestion or pollution occur on certain highway segments, a differentially higher congestion or pollution levy can be added to the fuel tax for those using the affected segment. This is difficult to administer in practice but if the area affected is sufficiently broad it can be implemented well enough to have a positive impact upon highway congestion and that portion of overall pollution caused by vehicular traffic.

In other words, a properly designed fuel tax supported by licence and registration fees for vehicles, trailers and other highway user types, can come tolerably close to what is meant by marginal social cost. Indeed, this tax itself is a kind of synthesized price for highway services. It is not of course easy to determine the economically correct level of taxation

nor the correct value of marginal cost for *each segment* of the overall highway system. There are many conceptual, empirical and analytical problems involved as indicated earlier. But acceptance of the principle of linking specific taxes to variable highway costs will itself provide incentives to improve and sustain such a linkage.

- (b) *Toll roads*: Various highway segments may be constructed, improved or simply designated as public or private toll roads administered by some authority. The function of the authority would be to recover the full costs of the system. Under either private or public ownership, the authority could be mandated to establish a set of prices which, with projected traffic of various kinds, would permit the enterprise to cover costs, including a "normal" rate of return, with receipts. This could take the form of Ramsey or constrained market pricing based on the inverse of the elasticity of demand for the several services or products that minimizes the loss of consumer surplus while providing normal profits and efficient production.¹⁵¹ This is especially important in situations where variable or marginal costs are declining, where large amounts of costs do not vary much with traffic volume or output and/or where economies of scale exist. In other words, if properly done, this represents an "optimal departure from marginal cost pricing" that avoids both subsidy and internal subsidization and permits strictly private operation and/or ownership. In this way, various highway segments could be more efficiently priced, and expanded or contracted, as cost and demand conditions warranted. At least, obviously "unprofitable" links in the system could be allowed to deteriorate rather than require ever more public resources to maintain even when traffic did not warrant.

Such authorities could of course exist only for selected corridors. The learning experience of attempting to link highway investment, finance and pricing together on a market-oriented basis, might well prove worthwhile. More and more activities hitherto viewed as having to be publicly owned and operated have now become privatized, in many cases with dramatic improvements in both efficiency and service quality. Such may be the case here.

- (c) *Highway trust funds*: The linkage between costs and outlays is formalized under a highway trust fund analogous to that in the United States. Federal taxes on fuel, parts and so on are allocated directly to the Highway Trust

Fund (established in 1956) and used for construction of federal aid highways although there is no pretense that all of the taxes are strictly user-charges as defined above.

Before 1956, highway expenditures in the U.S. had been part of general revenues into which receipts from highway user taxes had been placed, as is the case in Canada now. The trust fund arrangement allows federal aid to highways to be more carefully planned over a longer time. Since future proceeds from the designated highway taxes can be forecast with reasonable accuracy, highway programs are more definite and no longer have to engage in an annual competition with all other federal programs. Outlays for highway maintenance can be regularized and not interrupted from time to time because of budgetary considerations.

The linkage involved is more consistent with efficiency criteria as noted above and also with the principle that highway users and beneficiaries should pay for the costs they occasion. In addition, if higher user-charges are needed, they are likely to be more acceptable if the users know they will be spent on highways rather than on something else.

On the other hand, extension of the trust fund concept would seriously limit the discretion that the federal government has over its expenditure patterns and could impede effective fiscal policy or important reallocations of government outlays. The current fuss about holding excess revenues in the trust fund for airports and airways in the U.S. for reasons related to the overall budget deficit rather than responding to the needs of overcrowded skies is an example of one kind of problem. Although widely heralded when initiated and with a reasonably good performance, the Highway Trust Fund must also be viewed in the larger framework of overall fiscal policy where its virtues are less obvious.

7. LESSONS FROM THE PAST: RELEVANCE TO CANADA

It might appear that there are only negative messages here for any other country, explaining only how to avoid some of the more egregious mistakes. Even if there were some positive aspects, it is unclear whether another country would profit much from them. Wilson's "second law" is that "no country learns from the mistakes of another." That is however

rather defeatist and contrary to the purpose of this entire undertaking. Besides, even if the lessons are negative that does not make their message useless. Nor, given the proximity of Canada and the U.S. in ways well beyond geographical including the new forms brought about by the Free Trade Agreement, is it beyond the realm of possibility that some of the positive aspects of U.S. passenger transportation policies would be of some interest and even guidance to Canada.

Let us begin, however, with a certain paradox. Canada altered its transportation policy formally long before the U.S. and in a direction that the U.S. eventually copied. From erstwhile teacher to reluctant student and follower needs some explanation.¹⁵²

WHY CANADA WAS FIRST

The MacPherson Report of 1961 led to new legislation by 1967. Although some minor skirmishes occurred earlier in the U.S., as noted in Sections 4 and 5, it was not until the early 1960s that the first official reports began to emerge suggesting certain reforms but compromising on the need for regulation. There was President Kennedy's message but it went nowhere. The big lag in the U.S. was legislative — about 13 years, depending upon which modal reform measure used, compared with only six years in Canada.

There are several reasons for Canada's earlier success. The MacPherson Report was the result of investigations by a Royal Commission which gave it far more stature in Canada than any of the reports reviewed in this study have in the U.S. Congressional studies, presidential study groups, advisory commissions or task forces inevitably smack of partisanship and tend either to be very controversial or lacklustre — a kind of mixed bag to accommodate all the conflicting interests. As such they are far from being the kind of persuasive document that the MacPherson Report was, which focused strictly on the national interest. It not only emerged from a Royal Commission, which put it one step ahead of any of the U.S. transport studies in terms of prestige, but it was extremely well done and consistently argued. In short, it was persuasive in a sense that the U.S. reports were not.

The U.S. legislative process, even when a majority of legislators may be fully convinced of the desirability for some action, is inevitably more protracted for several reasons. During much of the period under review, the President

was from a party different from either the Senate or the House majorities or sometimes both. This leads to far more bickering, politicking and belated compromising than would otherwise be the case.

Transportation in the 1960s was more significant to the Canadian economy and polity than in the U.S. If it could be demonstrated that the system was operating under great handicaps by virtue of "obsolete" regulation and that greater reliance upon market forces would lead to large efficiency gains, the prospects of early legislative success would be enhanced. On the other hand, the more important the issues the longer the debates would be and conversely for relatively unimportant issues. This did not occur in the U.S., however, because transportation, while important, kept getting pushed down the legislative priority list. The greater involvement of the U.S. in world affairs, for example, led to a siphoning-off of Congressional and presidential concern. A trade-off between domestic and foreign concerns tended to predominate from the mid-1960s on including the Indochina War, the cold war, detente, and the more recent dramatic events in Eastern Europe and the ex-Soviet Union. Thus, not only was the case for deregulation not convincingly or forcefully presented as it was in Canada, it did not appear to require such urgent and immediate attention as other matters.

In addition, the U.S. regulation of transport, especially the trucking industry, was far more cumbersome, detailed and intrusive than that in Canada. It made the law and its interpretation more complex and, most importantly, it created a set of vested interests not only at odds amongst themselves but which resisted any change that affected them adversely or benefitted their rivals even relatively. The interest groups had powerful lobbies such as the Association of American Railways, the American Trucking Association, American Waterways Association, various shipper groups and labour associations representing each mode or broader labour unions such as the Teamsters, Longshoremen, etc. Any radical change in policy was bound to be extremely contentious thereby inviting Congressional rejection of controversial measures or emasculation of important aspects of particular bills. Even when legislation proposing somewhat greater reliance on market forces passed, it was frequently interpreted, by the ICC in particular, in such a way as to signal no change at all.

Finally, the process was prolonged by what some view as a greater propensity of U.S. business, labour organizations and private citizens to litigate and

protest actions viewed as contrary to their interests than in most other countries, Canada included. Since the issues were often complex and hypothetical¹⁵³ there was much scope for litigation. Indeed, frequently, the Department of Justice would oppose ICC rulings as did, later, the Department of Transportation. The government itself did not speak with a single voice on transportation issues.

These appear to be the most plausible reasons for the long U.S. delay. There may have concurrently been aspects of the Canadian polity between 1961 and 1967 that made reform legislation more expeditious than would otherwise have been expected. I leave this part of the story to others more informed.

CANADIAN-U.S. POLICY DIFFERENCES

There are, however, some aspects of both U.S. and Canadian transportation policies in general and passenger transport in particular that might usefully be raised at this point. The MacPherson Report relied upon intermodal competition in advocating regulatory reform and reliance upon market forces and said very little about intra-modal competition except where trucking was involved. Indeed, it relied upon the high degree of competition likely to prevail within trucking. It also relied on the possibility that own-account trucking and passenger car travel on increasingly ubiquitous, publicly provided highways would protect shippers and passengers from rail and air duopolies or tight oligopolies under deregulation.

The U.S., on the other hand, having sharply restricted competition within modes from 1887 through the 1930s, left intermodal competition to take care of itself. There were far more firms of each mode serving most city-pairs than existed in Canada. With freedom of entry there would have been even more firms competing — certainly enough to create workable competition and contestability. Yet progress in deregulating intra-modally was much slower than in Canada.

Nevertheless a paradox remains. While Canada relied largely, through necessity occasioned by relatively smaller markets and using the same transport technologies as the U.S., upon intermodal competition, there was never a policy of modal separation. On the other hand, the U.S., while restricting intra-modal competition up to the late 1970s, had a relatively strict policy against formal intermodal integration. In other words, the U.S.

was bent upon preserving the perceived specialized “inherent advantages” of each mode and competitive equality among them. From the vantage point of the 1930s and despite a much stronger antitrust tradition in the U.S., this seemed to require careful economic regulation and/or promotion of each mode including trucking and air transportation and greater reliance upon regulatory commissions than competition in transportation. Competitive forces intra-modally were to be constrained while those intermodally were to be encouraged — hence the twin policies of detailed intra-modal regulation and intermodal separation.

Canada’s prospects for intra-modal competition in rail and air were naturally more limited and the regulatory philosophy inherited from the U.K. less militantly antitrust. Canada viewed industry structure as a poor indicator of competitiveness in that a high degree of concentration probably reflected scale or at least financial economies and was thus to be tolerated but carefully watched and/or regulated to prevent abuses. With this viewpoint, Canada not only adopted a transport policy based on intermodal competition but permitted the railways to enter any other mode they chose and to any extent. The results in practice for Canada were not much different from those in the U.S. because the railways did little to integrate, say, truck transport into their operations but treated truck subsidiaries as separate and largely independent and autonomous profit centres rather than as instruments to stifle non-owned truck competition. Various public investigations of rail ownership of trucking companies since the 1967 Act have found that:

Nothing in this investigation, either the hearing phase of it or the information elicited by the extensive investigative work that we have done, reveals undue restriction of competition or prejudice otherwise to the public interest. . . . There has been this build-up of a trucking arm in CN and, as we have reported in this decision, Canadian Pacific is even larger in the trucking field than Canadian National. But these developments have proceeded alongside another development — the tremendous growth of that part of the Canadian trucking industry that is neither owned nor controlled by the railways.

The growth of the trucking industry and the fact that it is now No. 1 among all the Canadian domiciled carriers of freight in terms of domestic operating revenues generated — this and the moderate size of the foothold which the railways have achieved in “for-hire”

trucking — fatally flaws the contention that because the acquirer of Chalut is Canadian National Railways there is undue restriction of competition and prejudice otherwise to the public interest. That contention, certainly at this juncture in the development of Canadian transportation, is not well founded. Indeed, the investigation which we have conducted has exposed the other side of the coin: that the Canadian trucking industry at this time, and notwithstanding difficult problems that undoubtedly confront it, is continuing to forge ahead. The trucking industry is becoming of ever greater importance to the performance of the total transportation function in Canada.

A meaningful foothold has been gained by the railways in the trucking industry but ownership and control of their segment of the industry does not as yet approach a condition of restriction of competition, let alone “undue” restriction of competition.¹⁵⁴

Thus Canadian experience with permissive rail ownership of trucks and other modes has belied the fears that led the U.S. to stress modal separation. Not only has it had positive efficiency benefits in Canada but as Heaver concludes:¹⁵⁵

The railways have been able to participate in the growth of the trucking industry and have been able to perform their less than carload (LCL) common carrier obligations more efficiently than would otherwise have been the case. The existence of trucking subsidiaries allowed the rail operations to drop LCL services. Both the CN and CP hope that further integration of their express services with their trucking operations will lead to more efficient and profitable services.

The railway-owned trucking companies have provided leadership in Canada in efficient management practices, for example, in their concern for costing procedures. These advantages have become less evident as a number of large trucking companies with sophisticated and innovative management have developed. However, since a number of the large firms are foreign-owned, the development of the large Canadian and railway owned trucking companies may be perceived in Canada as providing the benefit of significant national ownership in an important industry.

These benefits have been realized while there is no evidence that railway ownership of trucking is having any adverse effects on the nation's transportation system. It is not leading toward the development of monopoly powers, nor is it likely to do so. It has not been associated at any time with the railroads discriminating in favour of their affiliated companies. There is no evidence that the availability of nationwide trucking and rail services is providing the railway owned companies with an advantage in negotiating with large shippers.

In sum, the Canadian experience with the railway ownership of trucking has been positive and is likely to continue to be so. Railway ownership of trucking is not now a significant issue in Canada.

This is another instance where Canadian policy has influenced that of the U.S. in transportation matters.

On the other hand, the Canadian concern for intra-modal competition was heightened by the airline deregulation in the U.S. in 1978, the rail and truck deregulation Acts of 1980 and their apparent benefits — some of which spilled over into Canada. In fact, these U.S. intra-modal initiatives were quickly followed in Canada by a new transportation framework entitled "Freedom to Move" dated July 1985 which formed the basis for the new *National Transportation Act* (NTA) in 1987 which, among other things, asserts that important economic objectives are likely to be achieved "when all carriers are able to compete, *both within and among the various modes of transportation* [emphasis added]" — a sharp change from the previous stress on intermodal competition. Other aspects of the NTA follow similar features embodied in the U.S. legislation.¹⁵⁶

In this sense, Canada was the first nation to embody "greater reliance upon competitive forces" into legislation in 1967 but the U.S. was first to deregulate intra-modal transportation.

It is now evident that, as far as the general policy of more reliance on market forces is concerned, both inter- and intra-modal competition is required. If competition is capable of being most effective where the firms involved have quite similar products and cost structures, as competition theory indicates, then it was folly from the outset in the U.S. to have attempted to thwart this, especially in air, truck and bus transport. Intermodal competition

where firms produce products of quite dissimilar characteristics and have vastly different cost structures, is apt to have specialized advantages for particular kinds of travel or traffic allowing spheres of modal monopoly power, without intra-modal competition therein. These were the so-called “inherent advantages” that the U.S. sought to preserve by its abridgment of the most effective competition, namely intra-modal. There was also much concern that the newer forms of transportation — air, truck and bus — in the 1920s and 1930s would need protection from rail predatory behaviour to develop their natural niches in the total transportation market. There was also concern that the regulated railways, suffering financially during the depression, needed some protection against the new modes not then subject to regulation and control. In this way “competitive equality” was joined with “preservation of inherent advantages” of each mode as rationales for the twin aspects of policy namely, modal separation and intra-modal regulation.

The problem with both of these was that they were based upon static views of both technology and inherent advantages. As the economy grew and changed, the relative importance of the quality elements became far more important than relative marginal cost differences among the modes for both travellers and shippers (higher values of time, comfort and convenience for passengers and just-in-time deliveries for shippers, for example). New technologies responded to these changes and also reduced the need for separate modal ownership and operations (for example, TOFC, COFC, RO-RO [Roll on-Roll off] etc.) as complementarities among the modes began to grow in importance, eroding the “inherent” differences which turned out to be less than immutable.

At the same time, the fear of rail predation against the other modes turned out to be vastly exaggerated. In all these experiences and trials, both countries learned from each other. Both intra- and inter-modal competition need to be free for maximum potential benefits.

Other differences between the U.S. and Canadian approaches to national transport policy may provide certain lessons for each country. On a very general level and subject to many caveats, Canadian policy is more centrist, deliberate and proactive whereas the U.S. is less centrist, more pragmatic and reactive.

In Canada, the federal government was closely bound up with railways from the start. In fact, construction of the railways was part of the deal for Confederation. In the U.S. the private sector was first involved, and the states and municipalities bribed, cajoled and otherwise sought to control rail development as much as possible through their territories. The U.S. federal government only later was involved in helping finance rail development.

In the U.S., central government acts mainly after it becomes clear that states or cities want it to on particular issues and/or when states cannot effectively act individually. Thus the *Act to Regulate Commerce, 1887* came after some states had already instituted railway regulation and were subsequently enjoined by the courts from regulating interstate commerce which was growing very rapidly. In Canada there was no doubt from the outset that rail regulation was a federal affair. Similarly in trucking, the U.S. regulated because the bulk of the business by the mid-1930s was interstate and the federal government had reluctantly become deeply involved in highway finance prior to the first highway Act in 1916. The Canadian federal government deliberately stayed out of interprovincial trucking regulation even though it was authorized to do it (Part 3 of the NTA of 1967) and left that authority to the provinces partly because of awareness of the messiness that similar regulation had caused in the U.S.

Of course, in both countries the role of the central government increased relatively since the early 1930s in transportation and elsewhere. In general, this creates national standards in such key areas as transport and health, and thus, there will be a tendency toward state, provincial and regional equality. This may not be very efficient however. For example, the U.S. interstate highway system was built to the same standards throughout the country regardless of actual or potential traffic. Nevertheless such national standards ensure that the poorest states and provinces will not fall progressively behind the more affluent in important aspects.

Both countries are now yielding to more state and provincial autonomy in part due to relative failures of economic policy at the federal levels, ideological trends to "conservatism," states-rights and the like, but largely due to huge federal deficits. If not pushed too far, this provides the opportunity for a variety of experimentation with policies in transportation and elsewhere that may yield valuable lessons more significant than forcing all into the mold of homogeneous national standards. To paraphrase President Kennedy, it may be beneficial to keep the nation safe for diversity.

In the final analysis, the greater apparent centrism and deliberateness in Canada with respect to transportation policy may not make much difference because the relative average size of each province to the total nation is many times larger in Canada than the U.S. (e.g. 10 provinces compared with 50 states). Thus, regardless of the degree of apparent centralization of power, the actual importance of the individual provinces in Canada on average is much greater than the individual state in the U.S.

LESSONS FROM THE U.S. EXPERIENCE

One is impressed that over longish periods of time the U.S. government has responded reasonably well to the dynamics of population growth and movement, technological changes and other shifts in the economy that have sometimes belatedly, sometimes prematurely been recognized. To be sure, few of the investments or subsidies have been thought of strictly or even casually in economic terms, but none of them have been totally unwarranted. Too much rail investment was probably stimulated by the land grants in the 19th century. Highways and airways have generally had excess capacity since the late 1920s. These may suffer from selected bottlenecks at present, due to increased use arising from inadequate user-charges and generalized economic and population growth. However, except for some more or less substantial rail disinvestment, abandonment of some early canals, building the interstate system with the same capacity throughout the country regardless of potential traffic in specific regions, and some airway and airport capacity problems, capital or public expenditures have not been wasted in transportation, despite the emphases upon defence, national cohesion and nation-building earlier. In some more recent cases, public reticence to expand capacity is more of a problem than capital extravagance.

It is in the realm of regulation that obviously uneconomic decision making has taken place in the U.S. transportation sector. Economic regulation of a particular industry by a separate, independent commission is something not to embark upon lightly. Even if the objective is to restrict excess profits and remove discrimination, and otherwise reduce the static welfare loss which itself is not the major waste involved, far more is required than rate of return and/or price regulation. The regulatory body must be prepared to control entry and exit, to monitor a structure of prices, to ascertain a reasonable proxy for marginal cost of specific services, to control service offerings and qualities, to monitor finances, to have some say in new

investments and acquisitions and, in fact, to become informed of and involved in virtually all aspects of the companies being regulated. It is not possible to select one or two variables and gear policy to these alone, expecting to make the performance of the industry more efficient without substantial costs to society both direct and indirect. Nor is it possible to stimulate what may be crucially important, namely the "entrepreneurial spirit." The "dead hand" of bureaucracy, political gamesmanship and the like are apt to divert attention from economic matters. Even single-firm monopoly cannot be regulated in terms of allowable rates of return without unwanted side effects such as inefficient production, expense padding and other aspects of the so-called A-J-W effect.¹⁵⁷ How much more complex it is to seek to regulate dozens or even hundreds and thousands of firms, as the ICC did following the *Motor Carrier Act of 1935*, with their varying cost and service characteristics, different capital intensities and other special features. The attempt to regulate such a hodge-podge is bound to fail if the objective is efficiency or the achievement of non-economic goals at least cost and the maintenance of some spirit of enterprise and risk-taking.

The experiment with industry-specific economic regulation has failed in the U.S. The message worth conveying to other countries is not to embark upon such a venture except in the context of safety. If the country is already regulating, it should begin the reform process with emphasis upon efficiency. Certainly it should not regulate to protect government-owned and operated railways. Canada does not need this lesson since federal regulation was never invoked for motor carriers. Provincial regulation, however, is not much better and should be reduced to at most a standard of "fit, willing and able" to new applicants offering a particular service.

In cases where, for political or less often, economic reasons, there is some propensity to regulate or regulations already exist, antitrust criteria should be applied. This, in turn suggests that "regulation" might best be left with the antitrust authorities, even in situations of true natural monopoly. Where antitrust or anti-combines laws are weakly enforced, as is the case in Canada, they should be strengthened and applied more frequently in such a manner that the business community becomes acutely aware of the consequences of violating the competitive standards. The consequences need to bear higher penalties than apply at present in Canada although the competitive implications of the Free Trade Agreement should be positive. If the antitrust laws are so weak or weakly enforced that they have little deterrent effect

upon anti-competitive business practices, it may be necessary for a specific industry to be directly constrained by a separate commission or board. If so, efficiency in the market sense should be the overwhelming criterion. The legislation should be precise, should give the board little discretion or opportunity to develop a "grand design," and its powers subject to a sunset rule after a decent interval in which an assessment of the performance of the industry is made and side effects evaluated.

Aside from the very few instances where the government owns and operates a productive enterprise which cannot be largely privatized, the rule should be that of a market economy. It should be monitored by strict antitrust laws vigorously enforced and should include a set of charges for use of publicly provided right-of-way and carefully applied effluent and congestion taxes so that social and private costs can be made to coincide as closely as possible. Under these conditions, competitive market forces will tend to bring about efficient outcomes, be responsive to consumer demands and flexibly adjust to changes in technologies, input prices, incomes, tastes and preferences. There will still be plenty for the government to do, but the microeconomy should be able to provide for the demands of the populace with a minimum of overt intervention but with continuous vigilance. The government can then devote more time to performing better on the macroeconomic level and concentrate on doing those things more appropriate to its capabilities.

THE MESSAGE FOR PASSENGER TRANSPORTATION

Canada led the U.S. in the general policy change area in transportation and indeed set a good example to follow. However, the U.S. led in a perverse direction with respect to rail intercity passenger transportation in creating Amtrak in 1971. Just six years later, and following the miserable example of the first five years of Amtrak's existence, Canada copied the U.S. "solution" to the rail passenger deficit problem. The results were pretty much predictable. Rail passenger service in both countries is largely uneconomic and its continued provision with subsidy jeopardizes the development of bus service in areas of low-density traffic and wherever both rail and bus competition exists. If some rail service is to be preserved, beyond the urban-suburban areas where it is essential for commuting purposes and where it is correctly viewed as part of the entire urban problem that goes far beyond mere transit, it will have to be confined within heavy-density corridors. In Canada this rules out any of the super high-speed alternatives with maximum

speeds approximating 300 mph and most of the alternatives in the range much above 125 mph as well because of lack of traffic density and the huge capital costs of the required exclusive rights-of-way.

As Soberman has suggested, higher speeds can be obtained on the Montreal-Toronto corridor in the 112–125 mph range which might offer some prospects of economic feasibility. But this would have to be accompanied by “improved reliability, and twentieth century passenger handling, ticketing and reservation systems.”¹⁵⁸ Some large investments in existing track, electrification along with tilt-train technology and improved communication systems would be necessary to achieve even this modest speed range. However, these investments might prove economically viable and the resultant three-hour service on the Montreal–Toronto run should be attractive. A similar situation exists in most of the U.S. corridors being considered although several of them may have sufficient density to warrant experimentation with the facilities designed for the higher speed ranges well above 125 mph, already achieved in regular service and higher still in experimental runs, over 300 mph.

The point to emphasize, however, is that maintaining rail intercity transportation options to provide access to remote areas or regions without any other public transportation alternatives is self defeating. Since the rail service is subsidized, this inhibits development of at least some unsubsidized bus alternative. It is not the sort of operation suitable for rail in the first place; it is not part of railroads’ “inherent advantage.” Clearly, this is an inappropriate reason for Amtrak’s or VIA’s existence although it is often given as a rationale. Whatever future they may have cannot be confined to such inherently uneconomic service for it jeopardizes a better and cheaper alternative, namely, bus transport. Since neither Amtrak nor VIA can become the nationwide systems they once were, high-density corridor business is all that remains where subsidy could be minimized if not eliminated. Even here, the prospects are bleak unless much more is invested in providing higher speeds.

The U.S. now requires the ICC to investigate bus owner complaints of predatory pricing by the subsidized Amtrak and to take steps to prevent it. That seems to be the least that should be done so as not to discourage more competitive development of the bus alternatives in various regions and for the specific clientele that has always typified bus travel. If subsidy

is to be given, its benefits would doubtless be larger, per dollar of subsidy, if provided for bus operations that were deemed necessary or desirable, if not efficient, to serve otherwise isolated regions or to provide at least some access for those unable to drive or afford their own vehicle.

"Some elements of the system clearly must disappear," Soberman concludes with respect to VIA Rail.¹⁵⁹ This applies to Amtrak too. It is merely a question of choosing which parts to eliminate, if not the entire operation, in the interests of a viable and competitive intercity *national* passenger system. The Doyle Report had it right even before Amtrak.

SOURCES OF COMPETITION IN INTERCITY PASSENGER SERVICE

If VIA or Amtrak are confined to corridor service, where abundant competitive forces both among and within modes already exist, what will ensure competition in the other segments of intercity passenger traffic? The private automobile provides much of the answer because of the ubiquity of the highway system. If rail service disappears from all but a few corridors, this leaves only bus and air service elsewhere. However, bus and air are scarcely competitive in the same markets. There is some complementarity but not much overlap. Bus transport will dominate over short distances and air over long. Each caters to a different clientele. Thus competitive forces in the lower-density intercity markets will rely primarily upon intra-modal rivalry. Some intermodal competition between air and bus may occur with the use of commuter-type planes or, perhaps later, tiltrotor aircraft. Further development of "intelligent vehicle-highway systems," linking "smart cars" with "smart highways" may significantly improve bus performance characteristics as well, (e.g. "smart" buses). This is highly speculative. But the future possibilities do hold out some prospects for enhanced competitiveness between these two modes for distances below say, 200 to 300 miles. Clientele may overlap depending upon relative costs. In the meantime, intra-modal competition will have to suffice.

Technically speaking, both modes have most of the features required for effective or workable competition. This was indeed the entire basis for deregulating the airlines in the first place. The same is true for the bus industry despite its penchant for high levels of aggregate concentration even when federal regulation was not overly restrictive because of benign neglect. It is because of the need to rely on intra-modal competition in each,

that the recent tendencies toward monopoly power in air and the perpetuation of concentration in the bus industry should be viewed with more concern. The problem in bus transport lies mostly with state or provincial regulation. Efforts to reform these constraints should be encouraged by the federal governments in both countries. The situation is more serious in air and will require strong remedial efforts already noted in Section 6. However there is nothing to suggest the need to return to the past cartel-type of arrangement in either industry. Nor are the prospects for more competitive behaviour unpromising. The view that competition is obviously workable in each industry should elicit appropriate responses to ensure that it is made to do so.

The "open-skies" movement and the Free Trade Agreement also provide prospects for increased competitiveness. Policy now needs to refocus explicitly upon the passenger side of the transportation market and concentrate again on air and, perhaps for the first time in any major sense, upon the bus industry. This Royal Commission itself is evidence of the new concern for passengers vis-à-vis freight and it is hoped that the emphasis will continue upon efficiency criteria à la MacPherson.

The regulatory reform process is not over.

There is a problem with respect to very low-density travel involving small communities which would not generate enough traffic on a regular basis to warrant either scheduled service or more than one carrier even with small, say 10 to 12 passenger buses, or commuter aircraft. The preferred solution will vary from case to case but efficiency cannot rely on competition among carriers eager to enter the market or actually offering service. Rather, service will need to be provided by non-profit organizations or contracted for through competitive bidding and financed with at least some public funds. In northern, remote regions of Canada where some rail services exist they may validly be continued even at a loss but the costs could probably be reduced "by eliminating VIA Rail as the middleman and having Transport Canada contract directly with the relevant railways."¹⁶⁰ Several alternatives are available that could sharply reduce costs even though overall profitability is unlikely.¹⁶¹

THE ISSUE OF SUBSIDY

General subsidies for multi-product enterprises serving diverse markets normally lead to inefficient pricing, especially in the contestable submarkets where below marginal cost prices create unfair competition for the

non-subsidized competitors. But passenger transportation services to remote, peripheral or sparsely populated and/or poor communities may not be profitable nor financially sustainable, except through subsidy. In such cases (and where the political authorities decide *some* level of service is needed in the public or social interest), an operating subsidy that makes up the difference between the revenues derivable from a specified service level and the avoidable costs under honest and efficient management would be warranted. Such subsidy will also be efficient as long as it is route-specific and the avoidable costs of the subsidized service are lower than any other mode or operator could achieve for comparable service levels. Contracting for such services is desirable along with close monitoring of the results.

Various subsidized route-specific services exist in Canada and the U.S., including in Saskatchewan, Pennsylvania and Maine, where experience has shown that close examination is necessary before specifying any management program for establishing additional subsidized services.

There is nothing inherently wrong with subsidies provided the goal they seek cannot be achieved more efficiently by any other means, including reliance on market processes. Too often societies have sought to benefit some group, such as poor or disadvantaged people in outlying regions, by non-specific subsidies, or general cross subsidization by large enterprises in the transportation industry. The costs associated with these subsidies go far beyond the amount paid, as they cause serious price and cost distortions elsewhere in the system. Furthermore, the so-called benefits do not meet the primary needs of the groups for which they are intended. The benefits of transport subsidy are essentially a system of partial income supplements tied to consumption of a specific good or service. These can be provided more efficiently to the recipients without causing distortions that could have wide ramifications within a particular industry. Income supplements, if made directly to recipients, preserve their freedom of choice. Cheaper transportation may have little impact on poverty levels or the quality of life of disadvantaged or poor people since the subsidy is confined to travel only which is often only a small subset of needs.

OTHER MESSAGES AND LESSONS

I began this essay by noting that three major themes had dominated the evolution of transportation policy in the U.S.: (1) Inconsistency and ad hocery, (2) Dominance of freight transport over passenger, and (3) Changing role of

economists.¹⁶² The policy changes since the late 1970s are not only the latest episode in these themes but they may have altered them in important respects.

- (1) *Inconsistency and ad hocery*: One of the major contributions of regulatory reform in transportation has been to raise the consciousness of the public to the superior organizing properties of competitive markets and the opportunity cost concept. This provided impetus for reform in other industries as well, including communications, banking, insurance, brokerage, power, gas and oil transmission, and others. Indeed, every area subject to some form of pricing and entry or exit constraints has been under increasing pressure to ease or eliminate them and to allow more discretionary and flexible decisions by individual companies. In short, regulatory policy virtually everywhere has adapted to competition by allowing more competitive-like behaviour and practices hitherto disallowed. Not all have been as successful as in transportation. One only needs to be reminded of the savings and loan debacle, although that was a case of botched deregulation. Most reforms have resulted in real savings to the consumer, greater efficiency and flexibility of firms to respond quickly to changing circumstances.

But the widespread understanding that efficient pricing and market orientation coalesce and that therefore much of the previous anti-competitive and anti-market regulations were not only superfluous but absolutely wasteful as well, means that one aspect of inconsistency and *ad hocery* in transportation policy in the past has been removed. Policy can no longer be completely open-ended as to purpose. Whatever the purpose, we now demand to know the cost and whether the purpose can best be achieved through anti-market manipulation of the transport industry or by some other less costly means. Even national defence is now expected to be cost effective! Whatever the purpose, stability, national cohesion, defence, etc., will not be pursued without an assessment of costs involved compared with alternative mechanisms.

Thus, for example, it is doubtful whether the *Act to Regulate Commerce* in 1887 could be enacted today. The evil sought to be remedied then was discrimination of various kinds. The first section of the Act declared that all rates be “just and reasonable” and the next three sections specified various kinds of discrimination that were unjust and unreasonable — essentially place, personal and long- and short-haul discrimination. The

present approach to such a situation, assuming the antitrust laws would not suffice, would be to increase competitive pressure by facilitating new entry of firms producing similar services. Of course, any private actions to fix prices and artificially exclude entry would be declared illegal as they are now under existing antitrust statutes. In no way would the current approach provide a small group of commissioners with power to cartelize such a huge industry in its entirety. Whatever solution were adopted today would be consistent with market forces not counter to them. Clearly, no commission would be allowed to attempt any grand design as did the ICC or to organize the scheduled airlines into a more effective oligopoly as did the CAB.

For Canada it is likewise apparent that the Crow's Nest Pass rates legislation would not be enacted again, not only because its stated purpose had long since been accomplished but also because of the mechanisms for railway redress recommended in the MacPherson Report and especially because there now exist more direct, more effective and less costly ways of accomplishing similar objectives. The same goes for the *Maritime Freight Rates Act*. Certainly, subsidy to passenger transport would not be allowed to piggyback, so to speak, upon freight revenues as already apparent in the creation of VIA and the general philosophy and emphasis on efficiency and market orientation. There are simply more effective and less costly ways of accomplishing the objectives of these three policies than were followed for many decades in Canada. National policy will henceforth not require higher real costs of transportation as the major instrumentality.

Thus transportation policy including investment policy can be expected to be more consistent in the sense that economic analysis will be given higher priority than before and efficiency criteria, "market conformability" and cost effectiveness will dominate. That does not mean that mistakes will not be made. Even proficiently handled benefit-cost analysis is no more than a guide to public investment and there are plenty of ambiguities in determining marginal costs, appropriate user charges and the like. However, policy should be expected to be based at least on reasonable economic analyses and data and less warped by such notions as the "grand transport mystique," the special status of being a "public utility" or a "natural monopoly" because of some alleged economies of scale properties and so on. With the present transportation options facing both travellers and shippers, there is no reason to single out the area for

treatment any different from other industries. Thus transport policy can avoid much of the ad hocery and inconsistency of the past. We can hope, for the sake of efficiency, it will do so.

- (2) *Dominance of freight transportation:* The uniqueness of the passenger business, as distinguished from freight, has now been assured by the separation of rail passenger service from goods movement. This is not yet complete but will be whenever Amtrak obtains more of its own right-of-way or is confined to that which it already has. Aside from this, there is no major transportation enterprise handling both large numbers of passengers and large amounts of freight in the U.S. Assets are largely dedicated to one or the other and operations and ownership for the most part are separate. For the first time this permits the inherent differences between the two industries to receive exclusive attention by all modes in the U.S. Furthermore, future transportation technologies such as maglev, high-speed rail, "smart" cars, tiltrotor aircraft are all passenger-oriented. While some are adaptable to freight (e.g. "smart" trucks), most are unique to passenger travel in the sense that they emphasize smoothness and comfort of the ride, as well as speed, dependability and safety which are arguably more important for passengers than for most freight. Freight transport technologies are likewise moving in their own more distinctive directions so that the differences between the two can be expected to widen. This will, of course, mean that any future policies toward transportation must specify either passenger or freight and not casually assume that because problems arise in freight transport their resolution should *ipso facto* apply to passengers, as has been the case in the past. As noted earlier, the creation of this Royal Commission is evidence that passenger service is no longer considered an unavoidable but unwanted stepchild of freight operation.¹⁶³ It is an important industry in its own right and sufficiently different from freight that even in rail transport it can no longer be viewed as a by-product.

If VIA is to continue, it will have to be given its own trackage over important route segments or at least over all routes it is expected to operate in the long run. The more complete separation of both ownership and operation from freight is essential for a valid test of whether there is any viable future for rail intercity passenger traffic in Canada and where that might be. If VIA cannot do it with ownership and operation rights and substantial freedom to price and market efficiently, then it should be closed down and its services leased to the highest bidder for one more

try. Part of the lesson of Amtrak implies substantial track ownership and control in the most important, heavy-density corridors and as complete a separation from freight operations as possible. Clearly any new higher speed passenger services will have dedicated and specialized track.

- (3) *The changing role of economists:* Since the transportation industries, including both freight and passenger, are now pretty much viewed as not fundamentally different from other industries producing either goods or services, microeconomists of all stripes can continue to study them without feeling guilty about invading the terrain of "transportation economists," as if that connoted some special expertise or knowledge about some unique type of economic activity. In fact, as noted earlier, the period when the economic case for regulatory reform was being fully developed and later implemented, involved the return of more general economic theorists to the study of transport problems. With the legislative successes now under their belts, there is some danger that microeconomists might drift away from transportation as they did after the late 1920s.

I doubt whether this is likely. More and more economists have had a taste of the area and have become involved with engineers and computer specialists in modeling various aspects of, for example, traveller choice, optimal routing, to say nothing of monitoring the consequences of the regulatory changes that have already taken place, the new patterns of price discrimination including Ramsey pricing, hub dominance, the net benefits already achieved and so on. There is much to do theoretically, empirically and in terms of systems-type modeling that intrigues most of the new generation of mathematically sophisticated and computer-literate economists.¹⁶⁴ Thus I see little prospect of neglect of the field merely because issues of regulation have more or less been resolved and part of the battle for improvement has been won. As Schumpeter long ago remarked: "specialists in applied fields, mostly in . . . transportation, had got further" than others in understanding price discrimination. He cites Hadley's, *Railroad Transportation*, 1886 as being the first to show that rate discrimination "may improve the situation of all parties concerned, including the one that is discriminated against."¹⁶⁵

The study of transportation has contributed much to economic theory. As Baumol and Bradford note, the "general line of argument (regarding optimal departures from marginal cost pricing) has appeared widely for

the better part of a century. The formal theorems date back more than forty years — this work has appeared in some of our leading journals under the authorship of some of the luminaries of our profession and was clearly not limited to a backwater of the literature.”¹⁶⁶

But if the study of transportation has contributed much to economic analysis the reverse is also true. Again it is Schumpeter who points out that “Any decent theory of cost and price ought to be able to make valuable contributions to railroad economics, and railroad economics ought to be able to repay the service by offering to general theory interesting special patterns and problems. . . . There are great possibilities in a co-operation of economists and engineers . . . few fields offer such possibilities as obviously as does the railroad business.”¹⁶⁷

It is just this close mutuality of interests and cross fertilization that ensures that transport will remain an active field of research for professional economists. Benign neglect is unlikely to recur even with the diminished interest in regulation.

This is true in Canada and indeed in most other countries. Some of Canada’s most prominent economists have devoted much time to transportation issues and continue to do so. The Canadian Transportation Research Forum, the Canadian Institute of Guided Ground Transport and many other groups of economists, engineers and computer scientists are actively engaged in transportation matters to say nothing of the many university institutes, majors and courses devoted to the subject. In this sense, Canada has nothing to learn from the U.S. If U.S. economists played a decisive role in the deregulation movement in the U.S., which demonstrably they did, it was mainly because such a highly visible role and attainment of high-level positions of administrative influence was necessary given the strength of the resistance to change compared with that in Canada. Both countries are thus well endowed with economists sufficiently concerned with transportation that the efficiency flag may be expected to keep flying for several decades at least or until some more grandiose paradigm displaces it.

ENDNOTES

1. Thomas K. McGraw, *Prophets of Regulation* (Cambridge, Mass.: Harvard University Press, 1984), p. 305.
2. Special Message to the Congress on Transportation by President Kennedy, The White House, April 5, 1962.
3. National Transportation Policy Study Commission, *National Transportation Policies Through the Year 2000* (Washington, D.C.: Government Printing Office, 1979).
4. *Ibid.*, p. 43.
5. For an elaboration, see George W. Wilson, *Economic Analysis of Intercity Freight Transportation* (Bloomington, Indiana: Indiana University Press, 1980), pp. 274–81.
6. *Royal Commission on Transportation* (The MacPherson Report) (Ottawa: Queen's Printer, 1961).
7. *Ibid.*, Vol. 1, p. 17.
8. *Ibid.*, p. 43ff.
9. For example, the ICC might hold rail rates well above those of barge lines to preserve the latter in case of military necessity. Such "protection" of a higher cost mode was called "umbrella rate making" and was one of the major problems tackled by *The Transportation Act of 1958*, discussed in section 4.
10. George W. Wilson, *Economic Analysis of Intercity Freight Transportation*, p. 280.
11. U.S. Department of Commerce, Office of the Federal Coordinator of Transportation, *Passenger Traffic Report* (Washington, D.C.: Government Printing Office, 1936).
12. M. Farris and H. Harding, *Passenger Transportation* (Englewood Cliffs, N.J.: Prentice-Hall, 1976), p. 241.
13. George W. Wilson, "The Relative Importance of Economic Regulation of Transportation vis-à-vis Everything Else," in *Economic Regulation: Essays in Honor of James R. Nelson*, edited by Kenneth D. Boyer and Willaim G. Shepard (East Lansing, Michigan: Michigan State University, Division of Research, 1981), pp. 24–25.
14. George W. Wilson, *Essays on Some Unsettled Questions in The Economics of Transportation* (Bloomington, Indiana: Foundation for Economic and Business Studies, Indiana University, 1962).
15. See George W. Wilson, "Economic Analysis of Transportation: A Twenty-Five Year Survey," *Transportation Journal*, Fall 1986.
16. Some have even gone so far as to translate passenger-miles into ton-miles by converting passenger weight, including baggage, into tonnage to obtain a presumed homogeneous unit of "stuff" to be moved various distances.
17. See the last issue of *Transportation Facts and Trends*, Transportation Association of America, December 1974.

18. Mark Hansen, "U.S. Intercity Passenger Transportation Policy: 1806–1990," in *Canadian Transportation Policy*, edited by David W. Gillen (Kingston, Ontario: John Deutsch Institute for the Study of Economic Policy, Queen's University, April 1990), p. 22.
19. Some states have bumper stickers that read, "Welcome to X. Please go home!"
20. See F. W. Taussig, "Railway Rates and Joint Costs Once More," *Quarterly Journal of Economics*, February 1913; F. W. Taussig and A. C. Pigou, *Railway Rates and Joint Costs*, *ibid.*, May 1913.
21. Cited in George W. Wilson, "Economic Analysis of Transportation: A Twenty-Five Year Survey," p. 35.
22. Alfred E. Kahn, Civil Aeronautics Board, 1974.
23. The pricing freedom was only partial in that it applied to traffic where the railways involved did not have "market dominance" and where they were deemed not to be "revenue adequate" overall. Needless to say the ICC attempted to construe these terms so narrowly that the rate freedom provisions of the *Staggers Rail Act* would have been largely nullified. In this, the Commission failed. In fact, it soon relented as new Commissioners were appointed or former Commissioners not replaced, leaving a majority in favour of the deregulatory initiatives made all the more pressing under President Reagan's broadside attacks on regulatory restraints of virtually any kind.
24. J. M. Munro and G. W. Wilson, *Road Transport: History and Economics*, Indiana Readings in Business, No. 35, (Bloomington, Indiana: Bureau of Business Research, School of Business, Indiana University, 1962), p. 55.
25. *Federal Aid Road Act, 1916.*
26. Mark Hansen, "U.S. Intercity Passenger Transportation," p. 27.
27. J. M. Keynes, *The Economic Consequences of the Peace* (New York: Harcourt, Brace and Howe, 1920), p. 9.
28. Thomas K. McGraw, *Prophets*, pp. 4–5.
29. See A. M. Milne and A. Laing, *The Obligation to Carry* (London: Institute of Transport, 1956), which is the classic study on the issue. For recent applications of the common carrier obligation see P.M. Shannon, Jr., "The Common Carrier Obligation in an Unregulated Environment," *Transportation Research Forum Proceedings*, 1980, pp. 476 ff.
30. I. L. Sharfman, *The Interstate Commerce Commission* (New York: The Commonwealth Fund, 1931), Vol. 3, p. 617–27.
31. Donald M. Itzkoff, *Off the Track, The Decline of the Intercity Passenger Train in the United States* (Westport, Connecticut: Greenwood Press, 1985), p. 30.
32. D. R. Owsram, "Icons and Albatrosses: Passenger Transportation as Policy and Symbol in Canada," in Volume 3 of this report.
33. Pennsylvania Truck Lines, Inc.–Control-Barker, 1 M.C.C. 101, 1936.
34. 79 Congressional Records 12206, 1935.

35. H. G. Moulton, "Fundamentals of National Transportation Policy," *American Economic Review*, December 1933, pp. 33–46 and *The American Transportation Problem* (Washington, D.C.: The Brookings Institution, 1933).
36. United States Senate, Special Study Group on Transportation Policies in the United States, *The Doyle Report: National Transportation Policy* (Washington, D.C.: Government Printing Office, 1961), pp. 221–22.
37. National Transportation Policy Study Commission, *National Transportation Policies Through the Year 2000*, p. 255.
38. *Royal Commission on Transportation*, (The MacPherson Report), Vol. II, p. 82.
39. Ari Hoogenboom and Olive Hoogenboom, *A History of the I.C.C.: From Panacea to Palliative* (New York: W. W. Norton, 1976), p. 110.
40. U.S. Department of Commerce, *Passenger Traffic Report*, p. 16.
41. Ibid.
42. Ibid., p. 7.
43. Ibid., pp. 1, 7, 26–28.
44. Donald M. Itzkoff, *Off the Track*, p. 13.
45. This is vividly portrayed in D. Itzkoff, *Off the Track*, pp. 15–18.
46. Ibid., p. 18.
47. George W. Wilson, "The Effect of Rate Regulation on Resource Allocation in Transportation," *American Economic Review*, May 1964.
48. For an elaboration of both "stories" see G. W. Wilson, *Economic Analysis of Intercity Freight Transportation*, chapters 1, 2 and 3. Referring strictly to demand see G. W. Wilson, "Notes on the Elasticity of Demand for Freight Transportation," *Transportation Journal*, Spring 1978, and references cited therein.
49. Cited in Munro and Wilson, *Road Transport*, p. 67.
50. *Historical Statistics of the United States*, (Washington, D.C.: Government Printing Office, 1975), Vol. II, p. 769.
51. George W. Hilton, *Amtrak* (Washington, D.C.: American Enterprise Institute for Public Policy Research, 1980), Table 1, pp. 3–4.
52. See Thorstein Veblen, *The Engineers and the Price System* (New York: The Viking Press, 1983), first published 1921.
53. Thomas K. McGraw, *Prophets*, p. 210.
54. See James M. Landis, *The Administrative Process* (New Haven: Yale University Press, 1938).
55. Hoogenboom, *A History of the I.C.C.*, p. 143.

56. The Hosmer Report, p. 72 (Reference No. 3, Table 1). See also George W. Hilton, "The Hosmer Report: A Decennial Evaluation," *I.C.C. Practitioners' Journal*, Vol. XXXV, No. 3, 1969, pp. 1472–73.
57. For a particularly illuminating comment on the efforts of the railroads at this time see D. Itzkoff, *Off the Track*, pp. 30–40.
58. George W. Hilton, *The Transportation Act of 1958* (Bloomington, Indiana: Indiana University Press, 1969), p. 13.
59. Hoogenboom, *A History of the I.C.C.*, p. 144.
60. Interstate Commerce Commission, *The Regulatory Issues of Today*, 1975, p. 2.
61. Cited in George W. Wilson, "Regulation, Public Policy, and Efficient Provision of Freight Transportation," *Transportation Journal*, Fall 1975, p. 8.
62. Interstate Commerce Commission, *First Annual Report of the Interstate Commerce Commission, 1887* (Washington, D.C., 1887), pp. 30–31.
63. Citations from Hoogenboom, *The History of the I.C.C.*, p. 150.
64. Jeremy Bentham (1748–1832) was the founder of utilitarianism in its more contemporary form. Briefly, this involves the attempt to measure the "pleasures" and "pains" or, in modern terms, the "benefits" and "costs" of any change in laws, regulations or specific decisions such as those made in every case before the ICC. The Benthamite notion was that one could somehow add up the benefits and costs of any action or decision and, if the former exceeded the latter, the decision was deemed good, desirable or a net improvement in welfare, and vice versa if the latter exceeded the former. Bentham hoped to be able to devise a metric, say units of utility and disutility, by which one might be able to aggregate and quantify the results of any decision — a kind of felicific calculus by which laws, decisions, etc., could be changed or reformed so as to contribute to the "greatest good of the greatest number or maximum net welfare or benefits." Although such a metric has yet to be devised and interpersonal utility comparisons are illegitimate, most economists have found that benefit-cost analysis of, say, public investments are extremely useful in making more rational decisions about them. Such analyses are direct offsprings of Bentham's attempt to create a more rational, positive and normative social science and indeed a basis for a rational moral, ethical and legal code. For a brief exposition, see Scott Gordon, "Utilitarianism," in *The history and philosophy of social science* (London: Routledge, 1991), pp. 248–70.
65. George W. Wilson, "The Effect of Rate Regulation on Resource Allocation in Transportation," *The American Economic Review*, May 1964, pp. 167–68.
66. U.S. Department of Commerce, *Modern Transport Policy; Documents Relating to the Report of the Presidential Advisory Committee on Transport Policy and Organization and Implementing Legislation* (Washington, D.C.: Government Printing Office, 1956). (The Weeks Report).
67. George W. Wilson, "The Weeks Report Revisited," *American Economic Review*, March 1959, pp. 130–31.
68. George W. Hilton, *The Transportation Act of 1958*, p. 204.

69. Interstate Commerce Commission, *Railroad Passenger Train Deficit*, Docket No. 31954, Decided May 18, 1959, p. 479.
70. *Ibid.*, p. 484.
71. George W. Hilton, *The Transportation Act of 1958*, p. 118.
72. Cited in *ibid.*, p. 119.
73. Interstate Commerce Commission, *Railroad Passenger Train Deficit*; Report prepared by Howard Hosmer, Hearing Examiner, et al., ICC Docket No. 31954, 1959. (The Hosmer Report).
74. Civic Aeronautics Board, Reports to Congress, *Annual Report, 1977* (Washington, D.C.: United States Government Printing Office, 1978), pp. 140–43.
75. For some data and discussion see Interstate Commerce Commission, *The Intercity Bus Industry* (Washington, D.C., May 1978), and Transportation Policy Associates, *Transportation in America*, 2nd ed. (Washington, D.C., 1984).
76. Interstate Commerce Commission, *The Intercity Bus Industry*, p. 45.
77. *Ibid.*, pp. 93–97.
78. The Weeks Report, pp. 1–3.
79. See for example, S. Huntington, "The Marasmus of the I.C.C.: The Commission, the Railroads, and the Public Interest," *Yale Law Journal*, April 1952; M. Bernstein, *Regulating Business by Independent Commission* (Princeton: Princeton University Press, 1955); and many others.
80. United States Senate, Special Study Group on Transportation Policies in the United States. *The Doyle Report: National Transportation Policy* (Washington, D.C.: Government Printing Office, 1961).
81. *Ibid.*, p. 38.
82. *Ibid.*, pp. 155–57.
83. *Ibid.*, p. 31.
84. *Ibid.*
85. *Ibid.*, p. 161.
86. *Ibid.*
87. *Ibid.*, pp. 434–35.
88. *Ibid.*, p. 322.
89. *Ibid.*
90. *Ibid.*, p. 323.

91. Ibid.
92. Ibid., p. 326.
93. Ibid.
94. Ibid.
95. Thomas K. McGraw, *Prophets*, p. 206.
96. The Doyle Report had this to say about multiple considerations that the ICC felt it had to examine (as noted earlier in Section 4 of this study) as an aspect of the "grand design": "Further, because there are so many factors to be taken into account by the I.C.C., from a national policy aspect as well as from the aspect of the relative importance of the conflicting evidence adduced in connection with any particular application, the I.C.C. can decide almost any case just about as it pleases and, by leaning upon, first, one aspect, and in another case, upon another aspect of our transportation policy, successfully withstand almost any scrutiny of the courts." (The Doyle Report, p. 124.)
97. U.S. Department of Transportation, U.S. Federal Railroad Administration, *Report on the Potential for Integrating Rail Service Provided by the National Railroad Passenger Corporation with Other Modes* (Washington, D.C.: Government Printing Office, 1976).
98. Hoogenboom, *The History of the I.C.C.*, p. 164.
99. U.S. Department of Transportation, U.S. Federal Railroad Administration, *Report on the Potential for Integrating Rail Service*, p. 2.
100. Ibid., p. 4.
101. Cited in Department of Commerce Memorandum, July 16, 1964 entitled "Summary of Legislative Activities Growing Out of the President's Transportation Message of 1962."
102. George W. Wilson, *Economic Analysis of Intercity Freight Transportation*, p. 272.
103. *Economic Report of the President*, January 1966, (Washington, D.C.: Government Printing Office), p. 128.
104. Bradley Behrman, "Civil Aeronautics Board," in *The Politics of Regulation*, ed. James Q. Wilson (New York: Basic Books Inc., 1980), p. 90.
105. For details and a methodological critique see G. W. Wilson, *Economic Analysis of Intercity Freight Transportation*, pp. 236–50.
106. Martha Derthick and Paul J. Quirk, *The Politics of Deregulation* (Washington, D.C.: The Brookings Institution, 1985) p. 38.
107. Ibid.
108. George W. Wilson, Memo, June 1966.
109. Robert C. Lieb, *Transportation*, 3rd ed., (Reston, Virginia: Reston Publishing Company, 1985), p. 47.

110. This story has been well detailed in several important sources, and will only be briefly related here. See McGraw, *Prophets*, chapter 7; Derthick and Quirk, *The Politics*; and James Q. Wilson, (ed.), *The Politics of Regulation* (New York: Basic Books, 1980), chapter 3.
111. Cited in McGraw, *Prophets*, p. 267.
112. A-J-W stands for the initials of the three authors, H. Averch and L. Johnson, "The Firm Under Regulatory Constraint," *American Economic Review*, December 1962, and S. Wellisz, "Regulation of Natural Gas Pipeline Companies: An Economic Analysis," *Journal of Political Economy*, February 1963. The "A-J-W effect" refers to the conclusion of these authors that rate of return regulation of a monopolist producing a single homogeneous product and seeking to maximize absolute profits will lead to resource misallocation in the sense that more capital, relative to labour, will be used than is needed for efficient production if the allowable rate of return is set between the opportunity cost of capital and the unconstrained maximum profit rate of return. A substantial literature grew up following these seminal articles in the early 1960s, most of which has little relevance to transportation regulation. For details see G. W. Wilson, *Economic Analysis of Intercity Freight Transportation*, pp. 186-98.
113. Alfred E. Kahn, *The Economics of Regulation* (New York: Wiley, 1970), 2 vols.
114. Passenger-mile data from U.S. Dept. of Commerce, Bureau of the Census, *Statistical Abstract of the United States, 1990*, 110th ed., (Washington, D.C.: Government Printing Office, 1990), p. 597: Data on number of trips refer to vacation trips only. See Steven A. Morrison and Clifford Winston, "The Demand for Intercity Passenger Transportation," *Transportation Research Forum, Proceedings*, 1983, pp. 526-34.
115. See U.S. *Bus Regulatory Reform Act*, U. S. Congress Public Law 96-261, 1982.
116. For details of these efforts see Edward L. Ramsdell and Imogene R. Burns, "Current Trends in the Health and Structure of the Intercity Bus Industry," *Transportation Research Forum, Proceedings*, Vol. xxvii, no. 1, 1986 and Clinton V. Oster and C. Kurt Zorn "Franchising as a Response to Increased Competition in the Intercity Bus Industry," *ibid.*, p. 59 ff.
117. Oster and Zorn, "Franchising," p. 60.
118. *Ibid.*, pp. 64-65.
119. See Itzkoff, *Off the Track*; Hilton, *Amtrak*; L. E. Tobey, "Costs, Benefits, and the Future of Amtrak," *Transportation Law Journal*, Vol. 15, 1987, pp. 245-302 and many others.
120. For details see Hilton, *Amtrak*, pp. 15-17, Tobey, "Cost Benefits," pp. 253-57.
121. U.S. Department of Transportation, U.S. Federal Railroad Administration, *Report on the Potential for Integrating Rail Service*, p. 39.
122. *Ibid.*
123. U.S. Public Law 94-280, 1976.
124. National Transportation Study Commission, *National Transportation Policies Through the Year 2000* (Washington, D.C.: Government Printing Office, 1979), p. 250.

125. A. E. Kahn, "Deregulation: Looking Backward and Looking Forward," *Yale Journal on Regulation*, Vol. 7, no. 2, Summer 1990, especially pp. 340 ff. Lieb, *Transportation*, Table 15-1, p. 317 shows a higher average number of carriers on various route types in July 1983 than July 1978: Adrangi, Gritta and Chow show reduced concentration levels for the industry as a whole by three major indicia of concentration for the post-deregulation period 1976-1985 in "Deregulation and Concentration in Air Transportation," *Transportation Research Forum: Proceedings*, Vol. xxvii, no. 1, 1986, p. 6 ff.
126. A. E. Kahn, "Deregulation," pp. 343-44.
127. *Ibid.*, p. 348.
128. *The Economist*, July 20, 1991, p. 82.
129. For a sympathetic and reasonable case see Tobey, "Cost Benefits."
130. Itzkoff, *Off the Track*, p. 131 citing the Executive Director of the Pennsylvania High Speed Intercity Rail Passenger Commission.
131. J. B. Hopkins, "Innovative Technology for Intercity Passenger Systems," in *Passenger Transportation in High-Density Corridors*, SPA-90-1. VNTSC, Cambridge, Mass., November 1990, p. 6.
132. M. Walrave and A. de Tessieres, "The French TGV System — Achievements To-Date and Future Developments," *World Conference on Transportation Research, Proceedings*, 1986, Vol. 2, pp. 1063-81.
133. U.S. Department of Transportation, "Assessment of the Potential for Magnetic Levitation Transportation Systems in the United States," June 1990, pp. 17-18.
134. *Ibid.*, pp. 301-02.
135. In Steven A. Morrison, "The Value of Amtrak," in *Canadian Transportation Policy*, edited by David Gillen, p. 104.
136. *Ibid.*, p. 110.
137. U.S. Department of Transportation, *Moving America, New Directions, New Opportunities* (Washington, D.C.: Government Printing Office, 1990).
138. *Ibid.*, p. 2.
139. *Ibid.*, p. vi.
140. *Ibid.*, p. 26.
141. *Ibid.*, p. 43.
142. *Ibid.*, p. 51.
143. *Ibid.*, p. 53.
144. *Ibid.*, p. 105.

145. Hansen, "U.S. Intercity Passenger Transportation," p. 21.
146. *New York Times*, June 20, 1991, p. 1.
147. Senator Daniel Patrick Moynihan, "Senate Overwhelmingly Passes Revolutionary Surface Transportation Act, 91-7," News Release, June 19, 1991, p. 1.
148. Fred P. Nix, "Road-user Costs: Report on Exploratory Research," *Journal of the Transportation Research Forum*, Vol. XXX, no. 1, 1989, pp 18–27; and Peter Bein, "Canadian Experience with HDM3 in Road Transportation Management," *ibid.*, pp. 9–18.
149. For details see Nix, "Road-user Costs," and George W. Wilson, *Economic Analysis of Intercity Freight Transportation*, pp. 91 ff.
150. Nix, "Road-user Costs," p. 26.
151. Interestingly enough the resurrection of Ramsey pricing was first in connection with efficient highway user-charges. See A. A. Walters, *The Economics of Road User Charges*, World Bank Staff Occasional Papers, no. 5, 1968, pp. 115–17. W. J. Baumol and D. F. Bradford, "Optimal Departures from Marginal Cost Pricing," *American Economic Review*, June 1970, facilitated the revival of Ramsey pricing.
152. This is not of course unique. The U.S. has viewed various aspects of Canadian policy as worthy of emulation. The most recent illustration is in the realm of medical and health insurance. But historically Canadians have seen themselves as, perhaps too often, followers of the policies and activities of the large, influential giant to the south. It is only in this sense that the transportation policy change in 1967 can be seen as paradoxical or at least one of a limited number of cases where Canada was well ahead of the United States in a major shift of public policy. The sequence is usually the other way around.
153. It could not be proved, for example, that greater reliance on market forces would improve things for shippers, travellers or many carriers. This was a theory about which no one could be absolutely certain. As such it was fair game for opponents to allege that those advocating the changes were "academics," most of whom had "never met a payroll" or misguided individuals with no personal stake in the outcome who did not mind experimenting with other peoples' interests.
154. Canadian Transport Commission, Motor Vehicle Transport Committee, *Report of Investigation and Decision, Pursuant to Section 27 of the National Transportation Act, in the matter of an objection by the Trucking Association of Quebec Inc. to the proposed acquisition of Chalut Transport (1974) Inc. by Canadian National Railways indirectly through its subsidiaries*, Decision No. MV-27-32 (M-76-2), June 1976, Appendix I.
155. Testimony before the Interstate Commerce Commission, Finance Docket No. 30500, "Norfolk Southern Corporation — Control — North American Van Lines, Inc.," Vol. III, filed August 1984, pp. 00149–00150.
156. For some of these see Trevor D. Heaven, "The Changing Role of Government Intervention in Canadian Transportation," Centre for Transportation Studies, University of British Columbia, April 1988 (mimeographed).
157. For an outline and elaboration see Wilson, *Economic Analysis of Intercity Freight Transportation*, pp. 187–207 where the A-J-W effect and other theories of regulatory consequences are discussed.

158. Richard Soberman, "Canadian Passenger Transportation Policy," in *Canadian Transportation Policy*, edited by Gillen, p.14.
159. Ibid., p. 17.
160. Ibid, p. 16.
161. See for example, Richard M. Soberman and Adil Cubukgil, "A Model for the Privatization of Rail Passenger Services in Canada," *Proceedings of the Canadian Transportation Research Forum*, May 1989, pp. 111–23.
162. See section 1.
163. Paraphrase of statement by Joseph B. Eastman in preface to U.S. Department of Commerce, *Passenger Traffic Report*.
164. See C. Winston, "Conceptual Developments in the Economics of Transportation: An Interpretive Survey," *Journal of Economic Literature*, March 1985 and Wilson, "Economic Analysis of Transportation: A Twenty-Five Year Survey," for examples along these lines.
165. Joseph A. Schumpeter, *History of Economic Analysis* (New York: Oxford University Press, 1954), p. 978.
166. William J. Baumol and David F. Bradford, "Optimal Departures from Marginal Cost Pricing," p. 277.
167. Schumpeter, *History*, p. 948.

THE ROLE OF EQUITY CONSIDERATIONS IN THE PROVISION AND PRICING OF PASSENGER TRANSPORTATION SERVICES

Robin Boadway*

January 1991

I. INTRODUCTION: EQUITY AS AN OBJECTIVE OF GOVERNMENT POLICY

A precondition for discussing the role of equity in passenger transport services is a clear understanding of the role of equity in more generally defined economic policy. Not only does this provide a normative basis for considering equity in this special context, but it also makes clear how the various instruments for pursuing equity are interdependent. The most general form in which equity objectives can be addressed is by monetary transfers based on some index of well-being or need. The traditional tax-transfer mechanism appropriately designed to take household and personal circumstances into account can be used for this purpose. The use of other narrower instruments, such as in-kind transfers, subsidies on particular goods or services, or regulations to target specific groups (each of which may be relevant for the case of passenger transport services), presumes that general taxes and transfers are in themselves inadequate. It is helpful to know precisely if and why this inadequacy exists.

We approach this general question in two steps. First, we consider the role of equity as an objective of government policy. This involves outlining the role of government in a market economy and seeing what this implies for

* Queen's University.

equity as an objective among other objectives of government, especially efficiency. Second, we consider the instruments that might be used to achieve equity objectives most effectively. Subsequently, we apply these principles to the particular case of passenger transport services, and draw some tentative policy conclusions.

The reader is forewarned that some of the discussion in the early sections is a bit abstract, though not difficult. This is not intended to obscure the issues, which ultimately require us to make rather practical judgments. Instead, the purpose of the discussion is to put the judgments into a broader context to avoid the temptation of making simplistic decisions whose broader implications are not fully understood.

It is important to stress at the outset that any policy conclusions can only be tentative because in the realm of equity, value judgments are unavoidable. Depending on the value judgments one is prepared to accept, the policy implications can differ dramatically. This will become especially clear in the first two sections of this paper. Indeed, this will be true of economic policy evaluation more generally. Except in rather special circumstances, the role of equity objectives cannot be ignored, and not all persons are likely to agree with the judgment made in any particular case. This is also likely a relevant consideration in the passenger transportation case as well.

1. THE OBJECTIVES OF GOVERNMENT ECONOMIC POLICY

The evaluation of economic policy involves comparing the allocation of resources under alternative policies and ranking them according to some criterion. Ideally, we would like to use a criterion which is as objective or scientific as possible. However, it turns out to be impossible to avoid making value judgments a part of the criterion. In what follows, we summarize how value judgments enter into economic policy evaluation according to modern welfare economics.

We begin with the bare minimum value judgments that are taken as being “maintained hypotheses” in much of the literature. These are the assumption of *individualism* and the *Pareto principle*.

Individualism means that what ultimately counts in evaluating policy is its effect on the well-being of individuals in society. This is opposed to some

organic view of society in which what counts are the effects of the policy on, say, institutions or on abstract principles which have a value in their own right (for example, freedom of speech, non-discrimination, etc.). The notion of individualism typically carries with it the assumption that individuals are the best judge of their own well-being, referred to as the property of *non-paternalism*. There are circumstances in which the principles of individualism and non-paternalism are difficult to sustain. For example: Should children and people with mental disabilities be treated non-paternalistically? Should households be treated differently than individuals? Fortunately, the answers to such questions are probably of secondary importance to us for transportation issues.

The Pareto principle holds that, if a policy change makes at least one person better off and no one worse off, it represents a social improvement. It seems to be a reasonable requirement, although there are circumstances in which it may be said to conflict with other widely held principles.¹

If policies could be ranked according to these two criteria alone, policy evaluation would be a relatively simple matter. Unfortunately, that will not be the case. Most policies violate the Pareto principle; that is, they make some persons better off and others worse off. Thus, something more than these two principles will be needed to rank policy alternatives. Economists who are not prepared to make further value judgments have attempted to avoid the issue in one of two ways. In some instances, they have attempted to extend the Pareto criterion to circumstances in which, although some persons are made better off and others are made worse off, the former could hypothetically compensate the latter and still be better off. Alternatively, they have simply ignored redistributive considerations and treated a dollar as being worth the same no matter to whom it accrues. Consider each in turn briefly.

According to the first criterion, if the gain to one person measured in monetary terms exceeds the loss to another, it seems reasonable to suppose that the gainer could compensate the loser and still be better off. If so, the *hypothetical compensation test* is said to be satisfied and a Pareto improvement should be potentially possible. This notion of a *potential Pareto improvement* has been taken by some to be synonymous with an efficiency gain. If we can come up with monetary measures of changes in individual well-being, it might seem that, by simply summing up the gains

and losses thus measured, we could say whether a policy change improves efficiency of the economy. (We return below to a discussion of how we can measure gains and losses in monetary terms.) However, this is not a satisfactory way to deal with the matter for several reasons:²

- i. The criterion of a potential Pareto improvement is itself ambiguous. Whether the hypothetical compensation test is satisfied or not depends upon the form which the compensation is imagined to take. In the literature, there are various assumptions about the nature of the compensation. The early literature based on Kaldor (1939) assumed that the compensation simply reallocated an existing bundle of goods without any change in production. Later analyses assumed that lump-sum income transfers were used and that the compensation therefore caused production changes to take place. Finally, the compensation might only be possible using distorting taxes and transfers, in which case different production responses could occur. For a given policy change being evaluated, whether or not a hypothetical compensation test is satisfied could depend upon which form of compensation is hypothesized. For example, if the gainers could make lump-sum transfers of income to the losers, the test might be satisfied, whereas if the compensation involved redistributing the given bundle of goods, or if transfers could only be made using distorting taxes, it might not.³
- ii. For any given form of compensation, the ranking of alternatives by the hypothetical compensation test is bound to be incomplete, and may well be contradictory depending upon the starting point. That is, in comparing the pre-policy and post-policy outcomes for any given policy change, the compensation test may not be satisfied in going from either one to the other, so the ranking is incomplete. Or, it may be satisfied for both.⁴
- iii. The compensation test is virtually impossible to implement from market data. For example, it is well known that aggregating welfare measures such as consumer surpluses over persons will not indicate whether or not the compensation test is satisfied.
- iv. Perhaps most important, if the compensation is not actually paid, a policy change will in fact make some persons better off and others worse off. The fact that compensation could have been paid will not constitute to many observers an unambiguous gain in social welfare. To evaluate it as a gain in social welfare requires trading off the gains to the gainers with

the loss to the losers. In other words, it involves an interpersonal comparison of welfare, something which the compensation test was designed to avoid.

Despite the unsatisfactory nature of the compensation criterion, many economists have nonetheless advocated, either explicitly or implicitly in their practice, the simple summing up of gains and losses in monetary terms to various persons without regard to whom they accrue.⁵ Various justifications may be made for this. One is an appeal to the separability of redistributive from efficiency concerns. As we discuss below, there are certain circumstances in which this is a perfectly defensible position. Indeed, part of our later discussion details when that is the case for passenger transport services. Basically, this position segregates all additional value judgments into a particular set of redistributive instruments, rather than avoiding value judgments entirely. Another justification for treating the dollar gains as having the same weight for all is that pure ignorance or lack of information may prevent one from desegregating the aggregate gains to persons of varying circumstances. Finally, some persons have argued that, over the long run, the chances of being a gainer and being a loser will roughly cancel out, so it is pointless to try to treat them differently. For many of the issues we address below, these latter two arguments clearly do not apply.

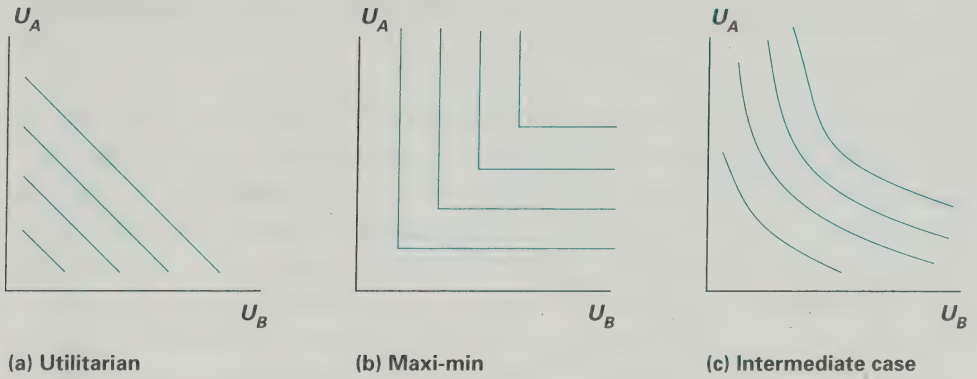
One is left with a fairly convincing case for going beyond the two maintained hypotheses of individualism and the Pareto principle for evaluating policies. The question then becomes what additional judgments need be made. There is bountiful and highly technical literature on *social welfare functions* which addresses precisely this issue. Much of it stems from the highly influential, but distressingly agnostic, work of Kenneth Arrow (1951). Arrow formulated the famous Arrow Possibility Theorem which, roughly speaking, says the following. Suppose we want to compare alternative allocations of resources and the only information we have available is the rankings of those alternatives by the persons in the economy. Suppose further that we place no restrictions on the form of the individual rankings, we require the Pareto principle to be satisfied, and we impose a technical requirement known as the *independence of irrelevant alternatives*. The latter basically says that the ranking of any two alternatives is independent of the availability of any other. Then, the only procedure which can be guaranteed to give a rational ranking of the alternatives is a ranking in accordance with the preferences of only one person, that is, what is referred to as a "dictatorship."

The Arrow Possibility Theorem has spawned a great deal of research, much of it attempting to relax its requirements so as to avoid the dictatorship outcome. A brief survey of some of the consequences of this literature is presented in the Appendix. For our purposes, we will simply avoid the conceptual issues involved in taking collective decisions by following the conventional procedure of assuming that society's ethical norms can be represented by a *social welfare function*. A social welfare function is simply a function which aggregates the utilities of the members of the society.⁶ In making a social welfare function operational for policy evaluation purposes, two main challenges emerge. The first is to obtain a measure of individual utility. Economists have devised a way of representing individual utility levels in monetary terms and this is referred to as a monetary measure of *real income*. The principles of representing utility by monetary measures of real income is also discussed briefly in the Appendix.

Given a monetary measure for utility levels, the next challenge is to devise a way to trade off or weight utility or real income levels obtained by different persons in order to aggregate them into a measure of social welfare. As is pointed out in the Appendix, when aggregating real incomes using a social welfare function, a key ethical property of the social welfare function is the degree of inequality aversion. Depending on the degree of inequality aversion, quite different degrees of redistribution can be called for. The three panels of Figure 1 depict indifference maps for three different social welfare functions with differing degrees of inequality aversion. All satisfy the properties listed above. In panel (a), the so-called *utilitarian* social welfare function is shown. According to this function, all that counts is total utility, not its distribution between the households.⁷ It is said to have no aversion to inequality. Panel (b) shows the opposite extreme of complete aversion to inequality. It is referred to as the *maxi-min* social welfare function.⁸ Panel (c) shows an intermediate case in which there is some aversion to inequality. These social welfare indifference curves can be used to rank alternative possible outcomes.

To illustrate the consequences of these differing attitudes towards unequal utility levels, consider the following example drawn from Sen (1973). Suppose there is a fixed amount of income to be redistributed between two persons, and suppose that the redistribution can be done in a costless way. That is, there is no inefficiency associated with the redistribution. Suppose that the two persons differ systematically in their utility functions. In

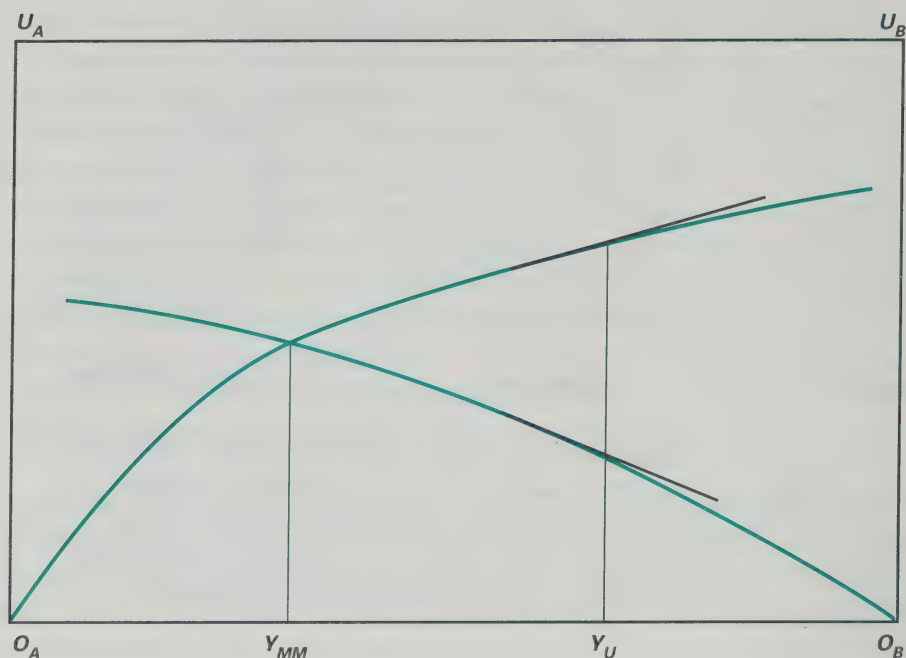
Figure 1



particular, person *A* obtains more utility from a given amount of income than does person *B*. We will say person *A* is the more efficient utility generator. It may be that *B* is disabled or less healthy. Consider now how the given amount of total income should be allocated between the two individuals under the various social welfare functions. Figure 2 can be used to do so.

Figure 2 depicts the utility levels associated with differing levels of income for the two persons, with *A*'s utility drawn from the origin labelled O_A and *B*'s drawn from that labelled O_B . Under the maxi-min social welfare function, the utility levels of the two persons would be equalized. The distribution of income Y_{MM} would achieve this. Notice that, given the differences in the ability to convert income to utility, this would require income to be distributed in favour of the less-efficient utility generator, *B*. This may be contrasted with the utilitarian case. Here, income is distributed between the two such that the marginal utilities are equal. This is shown as the distribution Y_U where the slopes of the utility functions are equal. Note that, in this case, the income distribution is opposite to the maxi-min case. Person *A* gets more than person *B*; also, person *A* ends up with a higher level of utility. The sum of utilities is maximized independent of its distribution. The intermediate case is not shown on the diagram, but obviously it will have a distribution of incomes between Y_{MM} and Y_U . More generally, the greater the aversion to inequality, the more income will be allocated to person *B* relative to person *A*. It should be obvious from this that value judgments play an important role in redistributive policy.

Figure 2



This example relied on differences in utility functions to generate ambiguities about the optimal redistributive policy. In those examples, if the two persons had identical utility functions, full equality would have been the prescription.⁹ However, it is easy to devise circumstances in which, even with identical utility functions, redistributive policy depends critically on the form of the social welfare function. This is discussed with reference to two further cases in the Appendix. There it is shown that simply making income variable by, say, allowing persons to decide their labour supplies, makes the extent of redistribution heavily dependent on the degree of inequality aversion in the social welfare function. This will be the case even when non-distorting (lump-sum) redistributive transfers are allowed.

When the tax-transfer system itself imposes efficiency costs on the economy, the amount of progressivity in the tax is reduced for any social welfare function. The examples in the Appendix clearly show that the extent of redistribution through the tax-transfer system is potentially quite limited when efficiency considerations are taken into account. I do not draw from that the conclusion that redistribution is necessarily limited; only that redistribution through the tax-transfer mechanism might be. As discussed below,

there are many other instruments that governments use for redistributive purposes. By their nature, many of them do not have the same efficiency costs involved. Instead, the extent of their use is determined by pure equity considerations.

A key issue that arises in evaluating and implementing economic policy is the extent to which the objectives of equity and efficiency can be separated. In an important general sense, the two cannot be separated. All policy decisions affecting resource allocation can, in principle, affect both, and, if government were a single monolithic agency, that agency would want to evaluate its decisions from a social welfare point of view incorporating both equity and efficiency considerations. However, the public sector itself is a highly decentralized organization in which resource allocation decisions are taken more or less independently in many different areas. In such a setting, it is more natural to ask whether each of these individual decisions should be taken with equity as well as efficiency objectives in mind.

Although the literature on this is not extensive, a good case can be made for continuing to follow the advice of Musgrave (1959) and think of the allocation and distribution branches of government as being essentially separate branches. Distribution decisions by their nature tend to be of somewhat broader applicability than allocation decisions. The latter are taken in the first instance at a more decentralized level. Most economists would agree that decentralized decision making on public sector resource allocation, including cost-benefit analysis, should basically be done with efficiency in mind. Equity objectives should be pursued by a central agency with a broader mandate. This agency could be responsible for setting tax-transfer policies and the regulatory framework, and these would act as constraints on those taking resource allocation decisions. Our discussion of equity considerations that follow are made from that perspective.

2. THE ROLE OF GOVERNMENT IN A MARKET ECONOMY

The above discussion focussed on the criteria that might be used to rank alternative allocations of resources, particularly those in which the distribution of utilities among households differed. We showed that very different rankings could occur depending on the value judgments one chose to make, in particular, depending on the degree of aversion to inequality in the social welfare function. Ultimately, we will consider the way in which

the government can achieve whatever redistributive objectives it wishes. However, before so doing, it is useful to review the prior question of the role of government in a market economy.

The conventional argument for government activity in allocating resources is a rather negative one, based on the notion of market failure. It is well known that, under certain circumstances, the decentralized decision making of the market has certain attractive properties in achieving efficiency in the allocation of resources. In particular, in idealized circumstances, the market, if operating competitively, will yield a Pareto efficient outcome. The role for government rests on two things — first, that those idealized circumstances may not hold (the market-failure argument), and, second, that the market outcome, no matter how efficient it is, may not be equitable. The ability of the government to intervene successfully depends upon how benevolent one views the government to be as well as how efficient it is (that is, how much *government failure* there is). Economists vary widely in their judgments of these issues, and therefore of how interventionist they think the government should be. This is over and above the differences that may exist in value judgments different persons hold about the ideal amount of equity versus efficiency. We leave that as an open question and simply recount the sources of market failure, thinking of them as necessary, though not necessarily sufficient, conditions for government intervention.

The sources of market failure and inequitable outcomes are fully documented in the public finance literature so we need do little more than list them here. They include the following:

Public Goods

Markets cannot be relied on to provide efficient amounts of goods which are simultaneously consumed by all. This is certainly true when the public goods exhibit the property of “non-excludability,” that is, when non-paying users cannot be excluded from use. However, it is generally also true when excludability is possible. Even if persons could be excluded from using a public good, efficiency suggests that they should not be since the property of publicness implies that additional users can be added at no cost to society. This is the traditional argument for government provision of goods and services.

Externalities

Related to the above is the case where activities of some firms or households cause beneficial or detrimental effects which do not get priced for one reason or another. Governments may respond by assuming responsibility for their provision, or for using corrective mechanisms such as taxes/subsidies or regulations.

Economies of Scale

The cost efficient scale of output of a good or service may be large relative to the market it is serving, in which case competition will not prevail. Again, government may undertake to provide the item itself, or it may regulate private behaviour. (Alternatively, it may do nothing as some economists would advocate.)

Unemployed Resources

There may be problems of coordination on some markets which imply that some resources go unused. Examples of this include labour, housing and capital. There have been some theories to suggest that there are systematic sources of inefficiency on these markets which government intervention can, in principle, address (for example, search externalities on labour and housing markets).

Absence of Full Markets in Uncertainty

A fully efficient allocation of resources requires that risks be fully traded and diversified away to the extent possible. This may not occur for a variety of reasons. For one, markets may be simply too thin for some types of risks, given the transaction costs involved. For another, a great deal of risk may be induced by uncertainty of government behaviour itself. Presumably this is not fully diversifiable. Finally, risk markets may not function perfectly because of informational problems as discussed in the next category.

Asymmetric Information

It is now widely acknowledged that many markets are characterized by what is referred to as asymmetric information problems, meaning that one side of the market is better informed than the other. This implies that resources will not be allocated in their most efficient way. The two most common

versions of this are *moral hazard* and *adverse selection*, and they have been analyzed most extensively in insurance markets. However, they apply to many sorts of markets, such as the markets for labour and capital. Basically, moral hazard refers to a situation in which one side of the market can take actions which affect the outcome, but which cannot be observed by the other side. Market outcomes may well involve non-optimal amounts of such actions. Adverse selection occurs when participants on one side of the market differ from each other in some characteristic that is not observable to the other side. Such markets are known to result in inefficient outcomes, and perhaps not even to have equilibrium outcomes. There is a considerable literature on market failure resulting from adverse selection and moral hazard. What has not been established in the literature is whether this sort of market failure can be averted by public sector intervention. For example, the public sector is likely to face exactly the same sorts of difficulties in becoming informed as the private sector. Thus, it may not be possible for the government to improve upon the market solution for such things as health and unemployment insurance, at least on efficiency grounds. We have to look elsewhere for a rationale for government intervention in these areas. The most likely reasons have to do with equity arguments which the final two items address.

Unequal Incomes

Even if the market were allocating resources in a perfectly efficient way, the outcome may not be regarded as being optimal, if it results in allocations which are unequal. Governments may therefore wish to interfere in order to redistribute resources on equity grounds. These last two categories look at equity in two related ways. One indicator of inequality is simply differences in observed incomes among households generated by the market. These differences may come from several sources including the following: abilities, inheritances, human capital accumulated, work effort and pure market luck. Virtually all governments engage in tax-transfer policies which redistribute income from the better-off to the worse-off households. However, there is a limit in the extent to which redistribution on the basis of income can be effective in achieving equity. There are two basic reasons for this, both of which have been referred to already. The first is that income may be a very imperfect indicator of well-being, especially since it does not reflect differences in, say, leisure. The second is that, for any given household, income is an endogenous concept. Persons can change their income by altering

their behaviour thereby causing inefficiency. The public finance literature has argued quite convincingly that the extent of redistribution that one can achieve by income-based instruments alone is likely to be quite limited.¹⁰

Social Insurance

There are, however, other arguments for redistribution which are related to the above but which recognize the possible usefulness of other sorts of instruments. The literature on redistributive income taxation tends to emphasize differences in the ability to earn income as a source of inequality. There are many other characteristics which also can result in utility differences and which can be (and, in practice, are) used as a basis for redistribution. They include such things as health and physical characteristics, employment status, location of residence and date of birth. Redistribution based on these features is sometimes referred to as social insurance for the following reason. The endowment each person brings with him or her at birth is largely a matter of luck. If individuals could insure themselves against being unlucky in these characteristics, they surely would. However, they obviously cannot do so on ordinary insurance markets. Insurance can only be purchased before the event being insured against is revealed and, in this case, that occurs at birth. Thus, they can only be “insured” against being unlucky at birth after the fact by the public sector. The conceptual device of putting oneself behind the “veil of ignorance,” that is, imagining not knowing what one’s characteristics are going to be, and asking what kind of insurance one would be willing to purchase, is often used as a normative justification for compensating persons for being unlucky at birth. This might be used as the justification for public health insurance, unemployment insurance, assistance to persons with disabilities, intergenerational transfers in favour of unlucky cohorts, etc. In each of these cases, the object of redistribution is somewhat different than income. Depending upon how sharply the government can observe the underlying characteristics, the possible induced inefficiencies in behaviour may be greater or less.

We take these roles for government in a market economy as given. The primary issues we wish to address concern how government achieves the equity objectives in practice, and how this may apply to the special case of passenger transportation services. The following two sections address those in turn.

II. INSTRUMENTS FOR ACHIEVING EQUITY IN THE ALLOCATION OF RESOURCES

Virtually everything governments do has a redistributive effect. What is perhaps less widely recognized is the extent to which redistribution is the main motivation behind various government programs. Public finance theory, which has tended to focus on efficiency analysis, has largely ignored this, at least until quite recently. We are used to thinking of the tax-transfer mechanism as the means of achieving distributive equity, with most other functions of government aimed at efficiency or stabilization issues. This way of looking at the role of government is typified by the public finance textbook device of separating government functions conceptually into allocation, distribution and stabilization branches. Government expenditures on goods and services are then thought of as fulfilling an allocative function, while distributive goals are met using taxes and transfers. However, closer inspection reveals that many important programs on the expenditure side of the budget are at least partly redistributive devices. This section simply summarizes all the instruments that fall into that category.

1. TAXES

Many of the taxes used by governments are structured to take distributional objectives into account. The main exceptions might be specific excises, the corporation income tax, and, to some extent, customs duties. Specific excises and customs duties are levied for particular purposes, such as correcting for externalities, industrial policy and user-fees. The corporate tax is best thought of as a withholding tax on behalf of the personal tax and, as such, its redistributive effects are addressed through the personal tax system by the method of integration (for example, the dividend tax credit). It also serves as a device for extracting revenues from foreign-owned firms by exploiting the international tax crediting system.¹¹ At the provincial and local levels, there are other tax types such as property taxes, licences and user-fees. These are also directed to non-distributive goals.

The two main taxes that are of relevance for redistribution are the individual income tax and general sales taxes (both federal and provincial). Each of these has provisions that are specifically redistributive in intent. Consider each in turn.

Individual Income Tax

There are three main ways in which the personal income tax is designed to take equity into account. The first is in measures which adjust the base for personal circumstances, such as number of dependents, disability and age.¹² The rationale for many of these measures is to make taxable income a better measure of welfare or discretionary income and thereby improve the horizontal equity of the tax system. As discussed earlier, this is justified by the notion that different persons require differing amounts of expenditures to achieve the same level of utility. As the Carter Royal Commission Report put it, different persons have differing levels of non-discretionary expenditures.

The second consists of deductions based on expenditures which are deemed not to be utility augmenting. These include the costs of employment, the cost of moving, educational expenses, day care expenses, charitable donations and medical expenses. These again are intended to address the horizontal equity issue by adjusting the base. The case of medical expenses, which includes those involved with disability, is an interesting one. It could be argued that persons who incur medical expenses and who receive the disability tax credit actually have different utility functions than others. Depending on the degree of inequality aversion one subscribes to, one might want to do more than simply make expenses for these persons tax-deductible. Doing the latter merely saves them the tax costs of the medical expenses they incur and does nothing to correct for the underlying disutility they live with as compared with healthy persons. If the degree of inequality aversion is high, one might want to transfer sizeable sums to persons with disabilities, especially since that is a characteristic which is relatively easily observable and not influenced by the behaviour of those involved.

The third component of the direct tax which addresses equity is the rate structure itself. Its degree of progressivity is a combination of the basic personal exemption and the structure of marginal tax rates on persons. In practice, it has been observed that the tax system as a whole has only limited progressivity in it.¹³ On theoretical grounds, there are good reasons for this. As mentioned earlier, the vast literature on the optimal progressivity of the income tax, emanating from the seminal paper by Mirrlees (1971), has tended to support the view that, given the equity-efficiency trade-offs involved, the amount of redistribution that should be accomplished through the income tax system is more limited than one might believe. The fact that leisure

and household production are sources of untaxed real income apparently restricts the optimal degree of progressivity considerably. Much of this literature is based on simulation procedures, and it would be beyond the scope of this paper to survey it fully. However, the results consistently tell the same message. We need not take from this the message that redistribution itself is of limited interest as a policy objective. As we point out below, there are many ways other than taxes that governments can redistribute, and some of them can be very effective.

Indirect Taxes

A high proportion of taxes are levied from indirect taxes of a general sort. These include the new *Goods and Services Tax* at the federal level and the provincial retail sales taxes, all of which have as their intended bases some measure of the consumption of households. The first thing to note about general sales taxes is that their bases could be designed to be virtually identical to direct tax bases, if the government so desired. That is, if the government levied a reasonably comprehensive income tax on individuals, it could design the sales tax to be equivalent to an income base. This would involve including all consumer goods and services as well as net capital goods purchases.¹⁴ Alternatively, if it taxed consumption at the personal level (which is not far from the truth under the current income tax system), the equivalent sales tax base could be purchases of goods and services used for consumption. In practice, of course, governments are somewhat inconsistent in their choice of direct and indirect tax bases, since they seem to prefer income for the former and consumption for the latter. However, the point is that they could, if they so desired, achieve the same base under either system.

This leads us to ask the fundamental question of the tax system. Why do we need both indirect and direct taxes in the system, if both can be levied on the same base? Why not just use one and save the separate cost of collection involved in having both? There are two answers to this — a theoretical one and a more practical one. The theoretical approach recognizes that there is a type of progressivity which can be achieved under a sales tax system which cannot be replicated under an income tax system. It arises because of the ability to implement a differential rate structure by commodities, such as charging lower tax rates (perhaps zero) on necessities and higher ones on luxuries. The main result of the theoretical literature is that,

if the utility function of households is characterized by *separability* of goods from leisure, nothing is gained by having differential commodity taxes alongside an income tax.¹⁵ To put it differently, the sales tax rate would be the same for all commodities and would be equivalent to a proportional income tax; thus, commodity taxation would be redundant with an income tax. Furthermore, even if separability did not apply, the government is unlikely to have the information to know the appropriate rate structure to use for the indirect tax system. Thus, the case for using an indirect tax alongside a direct tax on these grounds is weak.

The practical reason for having a direct-indirect tax mix lies elsewhere. The accepted reason is that evasion and avoidance of the income tax is possible, and the incentive to evade increases with the tax rate. Having a direct and an indirect tax in the same system thus fulfills two functions. First, it brings into the tax net on the expenditure side some persons who otherwise would have escaped taxation altogether because of evasion. Second, by providing a source of revenues to the government, it allows the tax rate on income to be lower, thereby reducing the incentive for evasion.

If we accept this as the practical reason for having a mix of direct and indirect taxes in the same system, then the case for introducing some progressivity into the system through the use of differential commodity tax rates can be made. Tax rates could justifiably be lower on goods with lower income elasticities of demand, if the possibility of evasion and avoidance is higher for high income persons. For example, the exemption for food and other necessities could be given theoretical justification. We return to this when we apply our discussion to passenger transportation.

2. TRANSFERS

Roughly one third of federal government program expenditures go to transfers to individuals. The most important of these are unemployment insurance, pensions and family allowances. At the provincial level, the major transfer program is for welfare. Welfare payments and family allowances are explicitly designed for redistributive purposes and complement the income tax system. Indeed, many have argued that they should be formally integrated with the income tax system in the form of a comprehensive negative income tax. This would rationalize the welfare system, treating it symmetrically with the income tax and avoid some of the existing system's adverse incentives on work effort.

The cases of unemployment insurance (UI) and public pensions are sometimes thought of as fulfilling other, non-redistributive, objectives. For example, the term “insurance” implies that UI is primarily a device for efficiency in labour markets. The term insurance applies to the notion of trading risk on the market. A purchaser of insurance is selling risk to an insurer who presumably is willing to take on the risk because of an ability to pool it among several purchasers. It is thus like a purchase of a commodity which both sides are willing to do voluntarily because it leads to a gain from trade, that is, to an efficiency gain. As long as there are no market failures, the private sector ought to be able to provide insurance efficiently. It is difficult, if not impossible, to make a convincing case based on efficiency or market failure grounds for public sector intervention in insurance markets. It is true that such markets may be plagued by such phenomena as adverse selection and moral hazard which prevent them from operating perfectly. However, the same informational problems which prevent the private sector from providing insurance efficiently apply equally to the public sector. This is as true for labour markets as for other markets. On theoretical grounds, it has been shown that private insurers could provide unemployment insurance almost as efficiently as the public sector.¹⁶ It is more reasonable to view unemployment insurance as being implemented at least partly for redistributive reasons, and that partly accounts for it being a compulsory scheme. There has been a limited literature on the use of unemployment insurance as a redistributive instrument, and it is likely to be an area for research in the coming years.

Similar issues arise in the case of public pensions. There is no particular reason to think that public pensions are justified as corrective measures for inefficient private pensions. Other arguments for their existence must be made. Three are found in the literature. The first is simply the *positive economics* argument that unfunded public pensions are transfers from the young to the old. The median voter obtains a positive net present value from such programs because, being somewhere around mid-life, the expected future benefits are well in excess of the contributions for the rest of his or her working life. Thus, the median voter will continue to vote for unfunded public pensions despite the fact that they reduce long-run levels of welfare for future generations.¹⁷

The second is that unfunded public pensions, being vehicles for redistributing income across age cohorts, can be used to smooth out fluctuations in “luck” among different cohorts. This is referred to as *intergenerational*

risk sharing and is sometimes used as a typical example of the principle of social insurance.¹⁸ As mentioned above, social insurance is the phenomenon of insuring *ex post* against the various circumstances of one's birth. It is distinguished from ordinary insurance which is intended to insure against future contingencies.

The third is a rather more recent argument. It views public pensions as a form of compulsory saving, or, in other words, a form of provision of future consumption greater than that which the person would have acquired individually. This can be viewed as a kind of *in-kind* transfer, much like the provision of food and housing. The recent theoretical literature on optimal income taxation focussed on the desirability of using in-kind transfers alongside income taxation as redistributive devices. The argument is that, given the optimal income tax, social welfare can be improved by forcing additional consumption of commodities that are relatively more important to low-income earners. The forced consumption, or in-kind transfer, will only be effective to the extent that the good in question cannot be retraded. It has been used in the literature to date mainly in the context of pensions, education and health services.¹⁹ This argument for pensions is really more applicable to compulsory funded pensions.

3. SOCIAL PROGRAMS

Governments in Canada also provide a number of social programs, many of which involve the provision of goods and services. Major examples of this include health care and services to the poor, those with disabilities and the elderly. We would argue that education at all levels can also be viewed as a sort of social program. One feature of these programs is that they support expenditures on what are essentially private goods and services. In the literature, they are referred to as *quasi-private* goods, since they are private but are provided in amounts fixed by the public sector. Being private, they could have been provided by the private sector. Indeed, in some countries, some of them are private, at least to some extent (for example, health care in the United States).

If one investigates why these quasi-private goods are provided publicly rather than being left to the private sector, the answer largely turns on their redistributive properties. There may be some efficiency arguments for public intervention revolving around externalities of various sorts, but they are

hardly enough to warrant the massive public intervention that we see in most industrial countries. It seems more appropriate to view public provision of health, education and welfare services as being instruments for redistribution. Indeed, from this point of view, they may accomplish at least as much as the tax-transfer system itself. Whereas the tax-transfer system tends to be restricted to redistributing according to income, these other programs, like UI and Old Age Security (and the Guaranteed Income Supplement), redistribute according to some other characteristic, such as health status or need. They can be viewed as types of social insurance which compensate persons for some characteristic which they have acquired independent of their economic activity. Many of these programs can be implemented effectively with adverse behavioural incentives that are not as severe as with the tax-transfer system based on income. This implies that the arsenal of redistributive instruments in the public sector budget includes taxes and transfers, UI, pensions, health care, education and welfare services. These span a major proportion of government fiscal activity.

4. FEDERAL-PROVINCIAL ISSUES

In Canada, a number of these activities are among the “exclusive” legislative responsibilities of the provincial governments. These include almost all health, education and welfare services, the exceptions being UI and pensions which are federal responsibilities by explicit constitutional amendment. This means that significant sources of redistributive power are in the hands of the provinces rather than the federal government. If left to their own devices, the provinces could, if they wished, pursue their own redistributive objectives. Two important issues arise here. The first is the extent to which the responsibility for equity ought to be at the federal rather than the provincial level of government. The second is, given that at least some responsibility lies with the federal government, how can it be exercised effectively, if so many of the instruments are in the hands of the provinces? Full treatment of these subjects is beyond the scope of this paper.²⁰ However, it is useful for background purposes to dwell briefly on these two points.

Regarding the assignment of equity responsibility to the two levels of government, the literature on fiscal federalism tends to support the notion that the federal government ought to be primarily responsible for equity. The main argument is the view that principles of horizontal and vertical equity ought to apply nationwide. In other words, identical persons ought to be

treated the same no matter where they reside. Conceptually, the social welfare function should apply nationwide. Against this grand principle must be set the counterargument that different provinces may have different “tastes” for redistribution. That is, some provinces may prefer to treat their poor more generously than others. The balance between these two arguments will determine where the ultimate responsibility for equity lies.

In Canada, the federal government assumes significant responsibility for equity and this is reflected in recent constitutional provisions. Specifically, section 36 of Schedule B of the *Constitution Act, 1982* reads:

- (1) Without altering the legislative authority of Parliament or of the provincial legislatures, or the rights of any of them with respect to the exercise of their legislative authority, Parliament and the legislatures, together with the government of Canada and the provincial governments, are committed to
 - (a) promoting equal opportunities for the well-being of Canadians;
 - (b) furthering economic development to reduce disparity in opportunities; and,
 - (c) providing essential public services of reasonable quality to all Canadians.
- (2) Parliament and the government of Canada are committed to the principle of making equalization payments to ensure that provincial governments have sufficient revenues to provide reasonably comparable levels of public services at reasonably comparable levels of taxation.

The first part of this seems to suggest that minimum national standards of equity should apply, and the federal and provincial governments have joint responsibilities for achieving them. The second part obliges the federal government to make equalizing transfers to the provinces. The *Canadian Charter of Rights and Freedoms*, which was also part of the *Constitution Act, 1982*, may also impose certain equity obligations or constraints upon the government, including non-discrimination, language and mobility rights. Thus, it would seem that the federal government has significant responsibilities in the area of equity, even though it does not control all the instruments for achieving it.

There are several components of federal-provincial fiscal arrangements that are designed to allow the federal government to play a major part in setting national standards of equity. One is the ability to maintain a common base and rate structure for the individual income tax. The vehicle for this is the Tax Collection Agreement negotiated with each province except Quebec. An agreement allows the federal government to administer the income tax of the province, provided the province abides by the base and rate structure of the federal government. The efficacy of the Tax Collection Agreement mechanism depends upon the federal government maintaining a dominant position in the income tax field. This has eroded significantly over the past few years as the provincial expenditure responsibilities have grown and the federal government has turned over to the provinces more responsibilities for financing them. The recent restrictions in Established Programs Financing (EPF) transfers and the advent of the *Goods and Services Tax* will accelerate that trend to the point where the harmonized income tax system is in jeopardy.²¹

Another vehicle for federal achievement of national equity standards is the system of equalizing transfers to the provinces. The main component of this is Equalization itself. However, the other two main programs, EPF and the Canada Assistance Plan (CAP), also have equalizing aspects to them. The ultimate aim of these programs is as stated in part 2 of Section 36, which is to provide provinces with the ability to provide comparable public services at comparable tax rates. The theoretical justification for this on equity grounds relies on the notion of horizontal equity. It is fully spelled out in Economic Council of Canada (1982). Note that the equalization provision provides for provinces having the capacity to deliver comparable public services at comparable tax rates. It does not oblige them to provide identical services. That is, some provincial responsibility for equity is maintained.

Finally, the use of the federal "spending power" via conditional grants can be seen, and justified, as a means of fulfilling the federal responsibility for equity. The sorts of conditions imposed by the *Canada Health Act* can be viewed largely as equity-motivated. The same might be said of the conditions of the Canada Assistance Plan. Historically, the use of the spending power has been a matter of contention from a purely legal point of view. The provisions of the *Constitution Act, 1982* would seem to have solidified the federal government's case for using the spending power as a necessary instrument (and the only one it has) for pursuing equity in areas of provincial legislative responsibility.

5. IN-KIND TRANSFERS

Governments may also provide transfers of goods to persons in need rather than delivering them through the income-transfer system. These are referred to as *in-kind* transfers. Examples include food (in the United States), housing, education, health care, welfare services and transport services to those with disabilities. There are two main reasons that are used to justify in-kind transfers as opposed to cash transfers. One argument views transfers to the poor as being at least partly motivated by the collective altruism of the better off. If the rich get utility from the well-being of the poor, they will willingly agree to make transfers to them. The efficient amount of such transfers may not come about through private charity because of a free-rider problem. That is, if all the rich simultaneously benefit from donations to the poor, it will not be in the interest of them individually to donate the optimal amount; there is no incentive for each one of them to take account of the benefit generated to others from a transfer. Thus, there is a role for public provision purely on efficiency grounds. If so, the ideal form of the transfer depends upon the preferences of the rich. If the rich cared only about the level of welfare of the poor, *as judged by the poor themselves*, cash transfers would be preferable to in-kind ones. However, if the rich would prefer the poor to adopt a different expenditure pattern, that is, if they were paternalistic, the transfers might be directed to certain goods. The one difficulty that arises here is that, if the goods can be resold (as in the case of food), the two types of transfers are essentially equivalent. Of course, it is also possible to have the transfer built into the general tax-transfer system through the system of deductions and exemptions. For example, owner-occupied housing is treated preferentially in the income tax system, while food is favoured under the sales tax.

In-kind transfers may also be justified on equity grounds as part of a more general system of redistributive policy towards the poor. The argument, due initially to Nichols and Zeckhauser (1982) and recounted in Blackorby (1990), is fairly technical, but the essential point can be summarized as follows. Cash transfers are typically related to the income level of households. Yet, income levels are very imperfect indicators of individual welfare because they do not capture other characteristics, such as health, employment, leisure, etc. Thus, a given income level can include persons of differing levels of welfare, and persons of high ability can behave like those of low ability so as to be eligible for transfers. In these circumstances, if

transfers could be targeted better, they could be directed toward those who truly need assistance at the expense of those who do not (but who can pose as if they do by varying their labour supply, etc.). One way of achieving this is to provide in-kind transfers in goods which are specifically related to non-income characteristics which affect utility levels. Others may include making transfers dependent upon the characteristic, and subsidizing the provision of the service in question. Different methods of targeting may be more efficient in different circumstances. Many of the above-mentioned services may be of this sort. More important for our purposes, transportation services for persons with disabilities and other disadvantaged groups would be of this sort. This will be discussed further in the next section.

It might also be noted that services provided for one group of persons may have spillover benefits for others. For example, passenger transport services to those with disabilities may benefit the able population by reducing delays which would otherwise occur from the disabled using the service. Thus, even if the main objective of providing the service is equity, efficiency considerations may support it as well.

6. REGULATION

Finally, governments may pursue equity objectives through the regulatory process thereby avoiding the budgetary process altogether. Examples of this include minimum wage legislation, pay and employment equity rules, and health and safety regulation both of products and the work place. The use of regulatory devices as an alternative to expenditure or tax measures will also be discussed further in the next section on equity in passenger transportation.

There is a general issue involved in regulating and providing in-kind services, and that is who ought to bear the cost. In principle, the answer to that is straightforward. The costs of any program instituted for equity purposes should be borne out of general revenues. There is no particular reason for the provider of the service to bear the cost. In practice, this may be difficult to enforce. It is not always clear precisely what the net cost of a regulation is to the sector being regulated. If the firm itself is in a regulated sector or is a public firm, it is even more difficult. Nonetheless, the principle is clear. To impose the cost on the firm is equivalent to taxing the users of the firm's services to pay for a policy whose objective is equity and which should be paid for by society at large.

III. APPLICATION TO PASSENGER TRANSPORT SERVICES

1. GENERAL PRINCIPLES

The above discussion evaluated the issue of equity in general and how it might be accomplished by policymakers. It was a rather lengthy discussion concentrating on rather abstract principles. This section is intended to apply those principles to the case of passenger transportation. In fact, the application is fairly straightforward and should take relatively little space. It should, of course, be remembered that the discussion can only lay out general qualitative arguments. The exact way in which they are applied in practice must necessarily involve a value judgment.

It is useful to begin this section with a summary review of some of the principles obtained in the earlier sections. Economists naturally think of the tax-transfer process operating on the basis of income as the main fiscal instrument for redistribution. There are many instances in the literature in which supplementary instruments can improve social welfare over and above what is possible by redistributive income taxation alone. The following summarizes the types of arguments involved.

- i. If the tax-transfer process is used to its optimal extent for redistributive purposes, and if persons differ in their ability to earn income, the only argument for using pricing policies, such as differential excise taxes or subsidies, for redistributive purposes is if the household utility function is not separable. In particular, goods which are relatively substitutable for leisure should be subsidized, and those which are complementary should be taxed.
- ii. If there are restrictions on the use of the income tax system for redistribution, a stronger case can be made for pursuing equity through pricing policies as a "second best" equity policy. Relevant restrictions might include the following:
 - a) The possibilities of tax evasion and avoidance induce virtually all countries to use a tax mix of indirect and direct taxes so as to reduce the benefits from evading and to indirectly tax those who do. If higher income persons are better able to evade, the structure of indirect taxation should be progressive.

- b) If the income tax system does not treat persons with negative taxable income symmetrically to those with positive taxable income, supplementary instruments which assist low income persons can be welfare improving.
 - c) The government may simply lack the information to be able to implement the optimal direct tax system. On the other hand, it may have good information about the sorts of commodities that are consumed by less well-off persons and treat them favourably.
- iii. If there are decreasing returns to scale industries, and if optimal taxes can be imposed, marginal cost pricing should apply (subject to the discussion of item ii above) and the losses should be recovered from general revenues.
- iv. Much of the literature on redistribution treats persons as differing only in income-earning ability. However, there may be other characteristics which cause utility differentials across households. If so, redistributive instruments which are based on these other characteristics can complement the tax-transfer mechanism in the government's arsenal of redistributive devices. Furthermore, the efficiency-equity trade-off might be much different here than in the income case. If the characteristics can be observed directly, the efficiency cost of redistribution along this dimension would be limited to the cost of raising the revenues required to finance it.
- v. In fact, governments use the expenditure side of the budget for purposes which have important distributional implications. Indeed, in many cases this may be their main justification. Uniform public provision of quasi-private goods through the public sector can be social-welfare improving if they have relatively greater benefits to those in the economy who are less well off. To the extent that these benefits can be earmarked to the less well off, they would be even more efficient redistributive devices. This is true even if the full potential of redistributive income taxes has been used.

The remainder of this section considers the application of these general principles to the case of passenger transportation.

2. ALTERNATIVE WAYS OF ADDRESSING EQUITY ISSUES IN PASSENGER TRANSPORTATION

It is useful at the outset to present a general catalogue of ways in which equity issues might be pursued in the provision of passenger transport services. This can be done via the tax system, by the use of subsidies or by regulation. A list of possible ways might include the following:

- i. Preferential sales tax rates may apply on passenger transportation services. They may be restricted to certain types of transport, for example, those which are used by those less well off. A case in point is the exemption of municipal transit from the federal *Goods and Services Tax*.
- ii. Related to this is the use of pricing policies on publicly owned transport, such as rail. Any price above marginal cost is essentially equivalent to an excise tax. There may be equity reasons for having differences between price and marginal cost, and there may be second best reasons as well.
- iii. Passenger transport services tend to have considerable infrastructure associated with them (roads, airports, rail lines, etc.) which gives rise to overhead costs which must be covered. This makes them essentially decreasing-cost industries perhaps over sizable levels of output, since the overhead costs must be spread. The public sector may provide or finance the infrastructure. From a welfare economics point of view, this actually makes some sense as a way of enabling the industry to set prices closer to marginal rather than average costs.
- iv. Special transport services might be provided on below-cost terms to the ill, the disabled and the elderly. These might include ambulances, wheelchair facilities and the like. They could be provided as part of publicly-supplied transport services, or their provision for private sector services could be subsidized or regulated.
- v. Transportation facilities serving remote and disadvantaged areas could be subsidized in the sense that they could be provided despite the fact that their provision would be rejected by a conventional cost-benefit analysis using consumer and producer surpluses.
- vi. Finally, there may be preferential prices given to identifiable types of persons such as the elderly, those with disabilities, etc. These could be

over and above the preferential prices that might be used as a method of ordinary price discrimination. Ideally, these should be financed out of general revenues.

This may only be a partial list in the sense that there may be other ways in which public policies could serve to interfere with pure market considerations in the provision of passenger transport facilities. However, these are illustrative enough of the types of policies that are used to serve our general purposes.

Some of the above involve the provision of certain services at less than cost, and an issue naturally arises as to how such services should be financed. As a matter of general principle, the ideal would be to finance (out of general revenues) any special costs arising from fulfilment of an equity objective. This might take the form of a subsidy or of public provision itself. However, that is not always entirely feasible. It may not always be possible to identify the separate costs of special facilities for such things as disability that may have been regulated by the public sector. There are certainly many examples of cases in which regulation of standards in other contexts is not accompanied by financial compensation, such as health and safety standards, building standards and zoning regulations. An alternative which might prove feasible in some instances is to allow a generous tax deduction or credit for additional expenditures incurred to satisfy transportation regulations which were designed to satisfy an equity objective. This will help pass some of the costs onto general revenues instead of relying on cross subsidization from other users.

3. SUBSIDIZING PASSENGER TRANSPORT AS AN INSTRUMENT OF INCOME REDISTRIBUTION

Suppose we think of passenger transport services simply as a good purchased by households along with several other goods. The amount purchased by households depends upon their income and preferences, and the government is able to observe income for redistributive purposes. Can we make a *prima facie* case for giving passenger transport services preferential tax treatment as part of a general policy of income redistribution? As mentioned above, there are two possible reasons why one might want to supplement redistributive direct taxation with preferential treatment by commodity. The first is as follows. Redistributive taxation taxes only the

real income earned from market activities; that obtained from non-market activities (that is, leisure and household production) go untaxed. If different commodities have different degrees of substitutability or complementarity for non-market activities (that is, if the utility function is not separable), a case can be made for differential commodity taxation. Unfortunately, the structure of that differential taxation is not straightforward, and involves both equity and efficiency effects. The equity effects would tend to favour lower taxes (higher subsidies) for necessities and higher taxes for luxuries.²² The efficiency effects would tend to impose lower taxes on commodities with the non-market activities, and vice versa. Thus, for example, a necessity which is also substitutable for non-market activities would seem unambiguously to call for a subsidy. However, conflicting cases could also arise.

The demand characteristics of passenger transport services presumably vary according to mode. One might think that bus transportation has a lower income elasticity of demand than rail, followed by automobile and air. The income elasticity for the bus might also be low relative to all commodities. If so, equity arguments might favour subsidizing the bus mode. If the transport services tend to be used to commute to work rather than for leisure activities, they will also be complementary with working, in which case the efficiency effects also tend to favour subsidizing them. If these sorts of empirical arguments are held with any degree of confidence, a theoretical case could be made that, even in a world in which progressive income taxes can be employed without constraint, it might still be justifiable to subsidize forms of transport relied upon by low income persons. On the other hand, other modes of transport such as intercity transport may tend to be more substitutable for working (that is, complementary with leisure) in which case one could propose a tax on efficiency grounds. Of course, there is the further problem that the government may not have enough information to design the optimal income tax. If the government nonetheless feels confident in knowing which types of goods are necessities, it can knowingly achieve some redistribution via subsidization of particular goods.

This general argument for subsidizing some sorts of transport services might be further supported by some other subsidiary arguments. One might be the well-known second best argument, which really has to do more with efficiency than with equity. The second best argument suggests that, if a commodity is priced below its marginal cost, other commodities which are substitutable for it should be as well. In the case of passenger

transport services, automobile travel is said to be priced below its social marginal cost because such things as the costs of congestion, road use and pollution are not fully included in its price.²³ Similarly, air travel may be priced below its marginal cost, if the full marginal-user cost of airports is not charged to the airlines. Since public surface transport may be to some extent substitutable for these modes, second best principles could be used to justify pricing below marginal cost.

The second argument relies on the notion that redistribution toward the poor may be partly justified by altruistic preferences of the well-to-do. If the altruism is paternalistic, it may support subsidizing particular types of goods. If transport services fell into this category, some support could be found for subsidizing them. It is not clear that the weight of this argument is particularly strong, compared with the previous two.

The above arguments considered the case for subsidizing some forms of passenger transport as a supplement to income-based tax-transfer mechanisms which could be used without restriction. If there are restrictions on the use of taxes and transfers, further support can be obtained for the subsidization of goods and services which are relatively more important to low-income persons. As mentioned above, perhaps the most important justification of having a mix of indirect and direct taxes is to counteract the massive evasion and avoidance that might arise if sole reliance were placed on direct taxes. That being so, the redistributive potential of direct taxes is not being exploited to its fullest. Under these circumstances, preferential tax treatment of necessities can be given theoretical justification. As well as food, shelter and clothing, transportation services used by lower-income persons could be given favourable treatment.

The direct tax system may, for some reason, not treat persons with negative tax liabilities symmetrically with those with positive tax liabilities. The former requires refundability of the liabilities and this may be difficult to implement fully. Also, some of the transfer mechanism is essentially in the hands of the provinces and they may not be fully coordinated with the federally-controlled income tax structure. Again, the preferential tax treatment of necessities may be a practical way of compensating for this absence of effective transfers to low-income persons.

4. SUBSIDIZING PASSENGER TRANSPORT AS A DECREASING COST INDUSTRY

Some types of passenger transport services may be characterized by decreasing returns to scale as a result of large fixed costs. In a perfect world, the fixed costs would be covered from general revenues in a non-distorting fashion, and the marginal cost would be imposed as a price for using the service. Capital decisions would have to be based on net surplus calculations rather than financial profitability. Any price charged above marginal cost should be regarded as equivalent to an indirect tax. The extent to which price diverges from marginal cost should be treated as equivalent to an indirect tax, and the above discussion applies. Thus, if it is desired for equity reasons to give preferential treatment to, say, passenger rail service, this service would have to be operated at a loss. In the case of buses, the recouping of fixed costs through taxes and fees of various sorts is apparently close to complete, as discussed in the Interim Report of this Royal Commission. Thus, in the absence of further indirect taxes, the practice is not far from average cost pricing, that is, above marginal cost pricing.

5. PASSENGER TRANSPORT SERVICES AS ASSISTANCE TO PERSONS WITH CHARACTERISTICS OTHER THAN INCOME

The above discussion concerns the incorporation of equity principles into the general tax and pricing treatment of passenger transport services. The emphasis was on redistributing among households according to income-earning ability and the possible shortcomings of the direct tax-transfer system. Now, we turn to the fact that there are other dimensions along which one may want to redistribute as well as the income dimension. It is obvious that, in practice, governments engage in many such measures. Recalling our earlier discussion, much redistribution takes place in ways other than the tax-transfer system and is based on criteria other than income. The health care system is a redistributive device based on health status; the pension system depends on age cohort; unemployment insurance depends upon employment status (and, to some extent, industry and regional location), etc. The question is whether there are certain aspects of passenger transport provision which can be viewed as contributing to equity according to non-income characteristics.

Incorporating equity considerations based on non-income factors is, in a sense, more ambiguous than those based on income. In the case of the latter, the equity concerns can be integrated into the income tax-transfer system, and the determination of the degree of progressivity can be left to those responsible for general income redistributive policy. With redistribution based on other factors, a somewhat independent judgment must be made on the extent of redistribution that is equitable. We saw earlier in our discussion of social welfare functions that, depending on the sort of judgment one makes about the degree of aversion to inequality, one can obtain very different optimal income distributions when persons differ in their utility functions. With very little aversion to utility inequality (for example, utilitarianism), income would be distributed in favour of persons who are more "efficient" at generating utility, and vice versa. For example, if persons with disabilities are assumed not to be able to convert income into utility easily, utilitarians would give them relatively few resources, that is, they would not compensate them for their disability. On the other hand, for social welfare functions with a great deal of aversion to utility inequality (for example, maxi-min), enough resources would be transferred to those with disabilities to compensate them for the difficulty of converting income into utility. This could involve substantial redistribution, especially if the efficiency costs of such transfers were low. What determines the outcome in this case is fundamentally a matter of value judgment, captured in what we have referred to as the degree of aversion to utility inequality. This makes it very difficult to state explicit policy implications since virtually any redistributive outcome can occur depending on the value judgment one makes. Even the direction of transfer of resources to persons with particular characteristics is a matter of judgment.

Furthermore, in evaluating this issue, it is not enough to know whether certain types of passenger transport services are associated with persons of certain characteristics who, for equity reasons, deserve special attention. It is also important to know if delivering the services through passenger transport is the most efficient way of doing so. If the characteristics can be identified and observed, it might be preferable to provide cash grants directly to the individuals rather than providing in-kind services. That, however, is essentially a matter of efficiency.

Potentially, there are a variety of characteristics affecting utility that might be associated with passenger transport services. In fact, a number of

them have been mentioned in the Research Program for the current Royal Commission. Let us treat each in turn.

Persons with Disabilities

Clearly one of the main disadvantages persons with disabilities face is a lack of mobility, so one would think that special provision of transport services might have a role to play here. To justify special treatment of persons with disabilities in general, it is necessary to assume that society has aversion to utility inequality. Suppose that to be the case. The assistance could take the form of cash transfers or in-kind services, or some combination of the two. There is now in place a system of cash assistance delivered to those with disabilities through the income tax. This includes a tax credit based on disability as well as the deductibility of additional medical expenses associated with the disability. One would suspect that, by any reasonable standards, the amounts involved in these two mechanisms fall far short of that which would be needed to compensate the disabled for the loss in utility due to their disability. Indeed, the implicit amount of inequality aversion contained in these measures is also probably a good deal less than that which is implicit in the system of redistribution based on differences in income-earning ability, unemployment status and illness. The tax credit as well does not distinguish among varying degrees of disability, as would presumably be required for redistribution truly based on differences in utility levels. One of the reasons for this is undoubtedly a difficulty in certifying with any degree of accuracy the relative degrees of disability of various sorts. In fact, there is apparently a good deal of uncertainty surrounding the exact criteria for eligibility for the Disability Tax Credit. Furthermore, the affording of deductibility for additional medical expenses represents only a limited attempt to compensate for the loss in utility due to disability. At the least, one might have expected full compensation to have been given for this, rather than the partial compensation implied by the tax savings from deductibility. At most, deductibility reflects the additional cost of earning income implied by the medical expenses.

In addition to the above forms of cash assistance, those with disabilities also obtain some services in kind. To the extent that their medical costs are insurable, these are provided free of charge.²⁴ Again, this compensates for many of the necessary medical expenses that are incurred as a result of being disabled, but does not attempt to compensate for the utility-reducing

effects of the disability itself (for example, pain and suffering, etc.). That is presumably the role of cash transfers. There are certain welfare services that provinces supply to the disabled that are analogous to full health insurance. Like health insurance, they represent transfers which are directly related to the extent of the disability. Persons who are more disabled, in the sense of requiring more medical and welfare services, receive correspondingly more assistance. This would seem to be an efficient way of differentiating among persons with differing degrees of disability.

The provision of transportation services to the disabled, including such things as wheelchair buses, special services at rail, air and bus terminals, specially designed seats, washroom facilities, etc., are directly analogous to in-kind medical and welfare services. They are a form of social insurance whose benefits are related to the degree of particular forms of disability. On equity grounds, it would seem to me to be arguable that these services should be freely provided to users for two reasons. First, that would put them on an equal footing with medical insurance, which itself must reflect an equity judgment. Second, from a value judgment point of view, it would seem to me to reflect a minimum level of aversion to utility inequality comparable to what seems to be accepted in the income tax-transfer system. As mentioned, there is an issue as to how these should be paid for. We have suggested that, in principle, the cost should be borne out of general revenues. This could take the form of a subsidy to the provider, in the case in which the provider is a private operator, though that may be cumbersome to administer. Perhaps the more sensible procedure would be to use a tax credit system for at least part of the costs involved in making the services available.

The Poor

The case for providing transportation services to the poor is quite different from that of the disabled, if by poor we mean those with low income. Of course, poor persons may be poor because of disabilities or, as discussed below, because of lack of employment opportunities in the province of residence. If so, the disability should be thought of as the cause. However, in general, we can think of poverty *per se* as reflecting differences in the ability to earn income rather than differences in the ability to transform income into utility. That being the case, transport services are merely one of the many types of goods that the poor consume along with a variety of others

such as food, housing, clothing, etc. Thus, as discussed above, policies for equitable treatment of these persons is primarily addressed through the tax-transfer system.

The case for providing in-kind transfers rests largely on an argument that passenger transport services are substitutable for non-market activities. It is not clear to me that such an argument can be sustained. Of course, if we have in place a large indirect tax system as well as a direct one, an argument can be made for incorporating equity norms into the tax structure. In this case, it can be held that services used relatively more heavily by low income persons should be given preferential tax treatment. This was discussed in more detail above so need not be repeated here.

Residents of Remote Areas

Residents of remote areas obviously seem to have greater need for transport services than other persons to obtain the same level of utility. Relatively large amounts of their budgets would be devoted to this as compared with other persons. Thus, it seems natural, at first sight, to think of providing such services at reduced cost. In judging this issue, a couple of considerations seem to be relevant. First, it is important to know whether the fact of residency in a remote area should be treated as an exogenously given characteristic of a person, or one which can be acquired at low cost. To say the same thing in other terms, one wants to know the extent to which the costs of moving from remote areas, *including* psychic costs, are high or low. If they are high, say because of cultural attachment to place of birth, then residency in a remote area might be treated as a non-income characteristic which directly affects utility. It then becomes a matter of comparing utility levels of persons who are resident in remote areas with other persons who are resident in populated areas.²⁵ If it is judged that residency in a remote area itself is a characteristic which causes utility to be lower, then a case could be made for compensating for that difference, assuming that one is averse to inequality differences. However, that compensation might well take the form of cash transfers, say, delivered as tax credits, rather than transfers in kind. In this case, the cash transfer need only be related to the fact of residency in remote areas and not other characteristics.

On the other hand, transfers in kind might be justified, if the degree of utility loss from living in remote areas was related to the level of consumption of particular services by persons. For example, the disutility of living in remote

areas might be related to the quantity of transport services consumed. Those with the greatest disutility might travel to and from the area most often, thereby incurring greater costs than others. It would not likely be feasible to make cash transfers to households based on their demand for travel services, since it would be difficult to measure the latter for any given individual. In these circumstances, targeting assistance to persons through subsidized transport services might be an efficient way of pursuing redistributive equity goals.

It might be noted that, if costs of migration to remote areas were not prohibitive, complications would arise. Residents of remote regions would contain a mix of persons who were there by birth, and who preferred to be there, and those who were born elsewhere and migrated there. The latter presumably require some financial compensation to overcome the psychic cost of moving to the remote region. In this case, the real income they obtain will differ systematically from their observed income, and this ought to be taken into account in the income tax-transfer system. In particular, the tax-transfer system based on observed income would systematically discriminate against moving to the remote region, since the differential income compensating for the psychic cost would be taxed. Thus, an inefficiently low amount of migration to the remote area would occur. The remedy for this would seem to be to provide preferential treatment through the income tax system rather than through in-kind services.

On the other hand, if mobility costs were low so that persons were free to move to and from remote regions, they would do so until they were in the region offering the greatest income. In this case, there would seem to be no apparent reason to differentiate through the income tax system between persons residing in different areas. In this case, whether in-kind services should be given special treatment would rely on the same general arguments as stated earlier about supplementing the income tax system with differentiated sales taxes. If transport services for persons in remote areas were substitutable for non-market activities, a case could be made for providing them at preferential cost. Also, in the indirect tax component of the tax system, if transport services were necessity goods to persons in remote areas, a case could be made for giving them preferential tax treatment in the sales tax system.

Residents of Have-Not Provinces

The case of residents in have-not provinces is somewhat similar to the case of persons in remote areas, though with some exceptions. The exception concerns the fact that the one important thing which differentiates persons in different provinces is the behaviour of their respective provincial governments. It has been well established that, in a decentralized federation where the provinces have significant taxing and spending responsibilities, there is an economic argument that can be made on both equity and efficiency grounds for a system of equalizing transfers.²⁶ The objective of such equalization is captured in the wording of subsection 36(2) of the *Constitution Act, 1982* which obliges the federal government to make equalizing transfers to the have-not provinces such that all provinces can provide comparable public services at comparable rates of taxation. This objective is realized in Canada by a variety of instruments including Equalization, Established Programs Financing and the Canada Assistance Plan. Equalization in principle works to compensate for differences in tax capacity among provinces.

The relevance of this provision for passenger transport depends upon, first, the extent to which passenger transport services are regarded as public services, and, second, on the extent to which they are regarded as the responsibility of the provinces. If they are regarded as public services, it will presumably be because of the sorts of equity reasons discussed above. To the extent that the provinces are responsible for their provision (for example, municipal transit, buses, roads, etc.), the federal government may, nonetheless, have an interest in setting equity standards, since by subsection 36(1) it bears joint responsibility for equity. In that case, the federal government can only exercise its influence over passenger transport by less direct means. For example, it may provide incentives through the tax system, or it may provide conditional transfers to the provinces to maintain some sorts of national standards. It cannot legislate directly in areas of provincial responsibility, and presumably this implies that it is unable to regulate standards. A full consideration of federal-provincial fiscal arrangements as they apply to passenger transport is beyond our scope. However, the ultimate equity objectives that the federal government might want to impose are the same as those we have already discussed.

The Elderly

Finally, let us consider one further category of persons not explicitly mentioned as disadvantaged, but nevertheless one that could be treated as such, and that is the elderly. This may be relevant since the elderly may have special reasons to rely on certain types of transport services. The case of the elderly is essentially the same as that of the disabled in terms of the principle involved.²⁷ Being elderly is a characteristic which is observable and which affects one's utility over and above that due to income level. One might want, therefore, to make cash transfers contingent on age for equity reasons. Also, to the extent that utility varies with need for transport services, provision of such services can be justified on equity grounds.

IV. SUMMARY REMARKS

The purpose of this paper has been to survey the arguments for using equity as a criterion for public policy; to discuss the ways in which equity considerations are incorporated into policy, including through direct and indirect taxes and transfers, social insurance, government expenditures, transfers in kind and regulation; and to consider what implications this might have for passenger transportation services. Some of the key results we have discussed are as follows:

- Public policy choices necessarily involve making value judgments. It is convenient to think conceptually of these judgments as being incorporated in a social welfare function, which is an aggregate of individual utilities. A key value judgment involved in this aggregation is the degree of aversion to inequality in utilities. Depending on one's aversion to inequality, very different redistributive policies can be obtained, from the progressive to the regressive.
- Individual utilities depend upon a number of factors, including ability to earn income, effort, and other characteristics such as health status and disability, age, employment and location of residence, to name a few. Redistributive policies can be directed at any of these dimensions.
- The conventional income-based tax-transfer system addresses mainly inequalities aimed at differences in ability to earn income. There is a limit to its redistributive potential, both because there are other sources

of utility than ability-to-earn-income differentials, and because income derives not only from ability to earn, but also from effort. In other words, there is an efficiency-equity trade-off. Studies have indicated that there are strict limits to the redistributive potential of income-based taxes and transfers.

- Governments engage in a number of other redistributive policies besides those based on income, and some of them are delivered through the income tax system. In fact, a significant proportion of program spending is actually motivated mainly by redistributive considerations, including such large programs as UI, pensions, medical care, welfare, and even public education. Many of them tend to be directed to characteristics other than income. In other cases, public provision of quasi-private goods can serve an income redistributive goal, if they are of relatively greater benefit to low-income persons.
- On the tax side, governments rely both on direct and indirect taxes to raise the bulk of their revenues even though, in principle, it might be better to rely on direct taxes only. Indirect taxes become necessary to reduce the ability and incentive to evade direct taxes. A case can be made for incorporating some progressivity into the indirect tax system.
- Applying these principles to the case of passenger transport, introducing equity considerations into the provision of passenger transport can be viewed as supplementing the income-based tax-transfer system, or as addressing utility differences arising from characteristics other than ability to earn income.
- As an instrument for income redistribution, subsidizing some forms of passenger transportation can be justified to incorporate equity into the indirect tax system. This would presumably apply to forms of transport which are used mainly by low-income persons. There are a variety of reasons discussed in the text as to why equity considerations ought to be incorporated into the indirect tax system.
- Subsidizing passenger transport could also be justified on grounds of efficiency. For one, these may be decreasing cost industries which would be operated at a loss under optimal pricing policies. For another, the second best argument states that, if some forms of transport are available below marginal social cost (for example, road transport), then others should be as well.

- Subsidizing of particular forms of passenger transport services can also be justified to provide assistance to persons on the basis of utility characteristics other than ability to earn income. This is particularly true of the elderly, the disabled and the ill of health, and it may apply to a certain extent to those in remote areas.

In summary, there are good theoretical reasons for taking equity considerations into account in the provision of passenger transportation. However, to do so requires that value judgments be made. It also requires some confidence in the relationship of various forms of passenger transport to individual characteristics which affect utility.

APPENDIX

The purpose of this Appendix is to discuss certain aspects of the literature on social choice and social welfare functions which have a bearing on the meaning and use of equity in evaluating public policy. The starting point is the Arrow Possibility Theorem which was mentioned in the text. Recall that it states that, if the only information we have available is the rankings of alternatives by the persons in the economy, if there are no restrictions on the form of those rankings, and, if both the Pareto principle and the independence of irrelevant alternatives must be satisfied, then majority voting procedures cannot be guaranteed to give a rational ranking of the alternatives.

In response to this Arrow Possibility Theorem, two branches of literature developed. One branch, known as *positive social choice*, is concerned with relaxing some of the technical requirements, such as the independence of irrelevant alternatives or the unlimited nature of household rankings, so as to avoid the dictatorial outcome. This literature is preoccupied with the mechanism of obtaining social rankings rather than the normative properties of the rankings, so is of little interest to us. The more interesting branch for our purposes is known as *normative social choice* analysis and looks at the consequences of using more information on household preferences. We restrict attention to that.

Under the Arrow Possibility Theorem, only household rankings or orderings of alternative allocations are allowed. Normative social choice theory considers adding, to that, information concerning the measurability and comparability of individual utilities. Naturally, this involves making value

judgments over and above those of individualism and the Pareto principle. The literature on normative social choice is much too long and complex for us to begin to survey. However, there is one notion that has come out of it that is very appealing for our purposes, and that is the following. Suppose we accept some notion of measurability and interpersonal comparability of household utilities. Sen (1977) has shown that, under very weak requirements, the social ranking of alternative social states depends only upon household utility levels achieved in those states, and not upon any extraneous information. The requirements are unrestricted domain, the Pareto indifference principle, and the independence of irrelevant alternatives. In the literature, this is referred to as the principle of *welfarism*. In other words, the abstract construct of a Bergson-Samuelson social welfare function can be used as a conceptual device for ranking social alternatives.

We take this notion of a social welfare function which depends only upon individual utilities as our starting point. Making it operational involves two further steps. One is devising ways of measuring individual utilities in, say, dollar terms. The other involves weighting these utility measures in the social welfare function. The latter naturally involves a value judgment. Let us discuss these very briefly in turn.

The conventional means for measuring the utility of a household is by the use of a so-called "money metric" indicator of utility. Technically speaking, a money metric associated with a given level of utility is the amount of money that would be required to achieve that utility level at a given set of reference prices for all goods (including leisure). This is what economists refer to as the *expenditure function*. Differences in the value of the expenditure function between two different allocations can be interpreted as the compensating variation or the equivalent variation depending on what set of reference prices is used. The conventional consumer surplus notion is simply an approximate measure of the compensating or equivalent variation. The fact that the value of the money metric depends upon the set of reference prices used is simply a reflection of the fact that utility cannot be unambiguously measured. It is useful to refer to this money metric indicator as the *real income* of households. If all persons had a common utility function, and if the same set of reference prices were used by all, the social welfare function could be rewritten as a function of these measures of real income rather than of utility levels. We return below to the form that the social welfare function might take.

Real income as measured above is a sensible money metric measure for utility if persons have the same utility function; that is, if they have the same ability to convert real income into utility. However, there are some obvious circumstances in which that will not be possible. Two that have been considered in the literature are as follows:

Differences in Household Size and Characteristics

When households consist of differing numbers of persons of different ages, they are likely also to differ in the utility they generate from a given level of per person real income. For example, some consumption goods are consumed in common among members of the household so that economies of scale can be achieved in consumption. Furthermore, some economies in household production of non-market services can be obtained. Economists have attempted to deal with these differences by adjusting real incomes by means of *household equivalence scales*. The household equivalence scale is constructed in the following manner. First, a reference household is chosen, for example, a single person with no children. Then, the equivalence scale for a person in a household of type x is calculated to be the ratio of the real income of that person to the real income that would be required by a person in the reference household to obtain the same level of utility. In other words, the equivalence scale can be interpreted as the number of persons in households of type x that are equivalent to one person in the reference household in terms of the ability to generate utility from real income. Equivalent real income is simply observed real income deflated by this equivalence scale. The concept of equivalence scales is widely used in empirical work for dealing with families of different size and composition. It could also be used to adjust real incomes to account for a variety of other circumstances, such as differences in the cost of earning income, medical expenditures, moving expenses, and the extra costs of living in remote areas.

It should be also noted that the concept of household equivalence scales bears a close resemblance to the principle of *horizontal equity* in tax theory. The latter says that persons at the same level of utility ought to be treated the same by the tax system (that is, ought to pay the same taxes). This idea is captured by the saying that “equals should be treated equally.” The implementation of horizontal equity should, in principle, be accomplished through the tax base. That is, the tax base should be adjusted by a system of deductions so that persons with equal utilities have the same tax base,

even though they have very different pre-tax incomes. The tax-rate structure can then be applied to the adjusted base. This procedure is generally used by economists as an argument for deductions rather than credits to account for such things as family size, medical expenditures, educational expenditures and the cost of earning income.

Differences in Utility Functions

A related problem is that of the case in which different persons with the same income receive different levels of utility from that income. For example, persons with disabilities or ill health may require more income to generate the same amount of utility. This is a more difficult one to deal with since there are obvious identification and measurement problems in correcting for these differences in utility functions. To correct for it fully in the tax system would presumably involve more than a set of deductions. Whether one actually wants to correct for it fully is discussed further below.

There are other ways in which utility functions may differ among individuals as well. For example, persons may have different preferences for particular types of goods. Thus, some persons may have a relative preference for leisure, others for fast cars, others for taking risks, etc. These differences would also be very difficult to take full account of because of obvious measurement problems. On the other hand, from the point of view of redistributive equity, it is not obvious that they are essential problems.

Suppose we have solved these problems of how to measure utility using some notion of real income suitably corrected for differences in household circumstances and utility functions. The next step in principle is devising a weighting system for aggregating these utility measures, or, in other words, devising a social welfare function. This is where the crucial value judgments arise. Depending upon the type of social welfare function chosen, very different redistributive policies can emerge as is shown in the text. If we are prepared to make further value judgments, a simple general form of social welfare function emerges which allows us to isolate a key determinant of redistributive equity.

To begin with, let us postulate some reasonable properties that a social welfare function might be required to satisfy. It will be assumed to satisfy the Pareto principle and to be welfaristic (that is, to depend only on utility

levels of households, and to be increasing in individual utility levels). Following much of the literature, the social welfare function will be assumed to satisfy the following further reasonable technical properties: anonymity, separability and quasi-concavity. Anonymity means that it does not matter which households get which utility levels; only the utility levels themselves count. Separability means that, in ranking two allocations, only those households which have strict preferences over the two should count, not those which are indifferent. (This was due to Fleming (1952) who used it as a basis for arguing in favour of a utilitarian social welfare function.) Quasi-concavity is a technical term which is equivalent to saying that social indifference curves in utility space cannot be concave to the origin. These requirements leave a large family of social welfare functions. However, they differ from one another in what may be termed their *aversion to inequality*. The notion of aversion to inequality is illustrated in the text for the simple two-person case. Here we may simply note the general algebraic form that the social welfare function may take:

$$W = \sum_{h=1}^H \frac{(u_h)^{1-\rho}}{1-\rho}$$

- where u_h is the real income of household h , H is the number of households, and ρ is the aversion to inequality, sometimes also referred to as the equity parameter. Note that for $\rho=0$, we have the classical utilitarian case considered in the text. Similarly, as ρ approaches infinity, the social welfare function approaches the maxi-min form, also discussed in the text. More generally, the higher the value of ρ , the greater the aversion to inequality in utilities among households, and the stronger the equity-efficiency trade-off.

The consequences of differences in the degree of inequality aversion for redistributive policy are illustrated in the text with reference to lump-sum redistributive transfers when total incomes are fixed. However, there are other circumstances in which the parameter becomes important as well. Two further cases will illustrate this. The first case is that in which we allow labour to be variable, but retain the assumption that income can be redistributed in a non-distorting way. The second occurs when we allow for the efficiency costs of redistributive policy.

Suppose we continue the above two-person example, but assume that the persons have the same utility function. Furthermore, assume that they have different wage rates because of some underlying differences in ability, and that they can vary their income by changing their supplies of labour. Thus, utility is assumed to depend upon consumption (after-tax income) and leisure. *A* is assumed to be the high-wage person and *B* the low-wage person. For this case, the government is assumed to be able to redistribute income in a lump-sum way, and we ask how redistributive the tax system should be. Consider the two extremes of utilitarianism and maxi-min.

In the utilitarian case, the government redistributes income until the marginal utility of consumption is the same for both persons. As Stiglitz (1987) has shown, equalizing the marginal utility of income entails a highly redistributive policy which actually makes the high-wage person worse off than the low-wage person. In the special case in which the utility function is the sum of the utility of consumption and the utility of leisure, full equality of income is achieved, but high-wage persons work more because of their higher wage rate.

The maxi-min case has been considered by Sadka (1976). Here, utility levels are equalized by the tax. Since the high-income persons will be supplying more labour, they will have to be left with high levels of income to keep them at the same utility level as the low-wage persons. Consequently, the tax system will be less progressive than under utilitarianism. In fact, there is no presumption that it will be progressive at all. Depending on the shape of utility functions (that is, the substitutability of leisure for consumption), the tax could be regressive or progressive.

Thus, even without bringing efficiency considerations into the picture, there is considerable ambiguity about how redistributive the tax system should be. It depends upon whether or not utility functions differ; whether leisure is variable and, if so, how substitutable it is for consumption; and how much inequality aversion there is in the social welfare function.

Once efficiency considerations are added, the analysis becomes more complicated, but the ambiguities remain. The complications arise because of the fact that redistributive taxation is no longer lump-sum, but is based on income. That being the case, the tax imposes a distortion on labour supply, and the distortion increases with the extent of redistribution. There is thus

a trade-off between equity and efficiency. Unfortunately, the analysis of this problem is very complicated, and most authors have had to resort to simulation techniques. There is an enormous literature on the so-called “optimal income tax” which calculates the form of the income tax schedule under a variety of supposedly realistic assumptions about individual preferences and social welfare functions. Basically, the message of that literature is that the optimal amount of progressivity of the income tax is surprisingly limited.²⁸ It is not clear from the literature how much of that is due to the efficiency-equity trade-off and how much is due to the variability of leisure *per se*.

ENDNOTES

1. See, for example, Sen, 1970.
2. A more technical and detailed review of these arguments can be found in Blackorby and Donaldson, 1990.
3. A full discussion of this issue may be found in Boadway and Bruce, 1984.
4. The former arises because some changes may involve redistribution with no change in efficiency. The hypothetical compensation test will be satisfied only if all persons could be made potentially better off by a change. If a policy has mainly redistributive effects, that will not be possible. For example, to use the technical jargon of welfare economics, a policy change which involved a move along the society's *utility possibility frontier*, making one group of persons better off and another worse off without distorting the economy, could not be ranked by a hypothetical compensation test. The test would not be satisfied either for the policy change or for the change in the opposite direction. The possibility that the compensation test is contradictory in the sense that it could be satisfied in both directions is well known and is referred to as the Scitovsky Paradox after Scitovsky, 1941. It is discussed fully in Boadway and Bruce, 1984.
5. One of the main advocates of this has been Harberger, 1971.
6. This is sometimes referred to as a Bergson-Samuelson social welfare function after Bergson, 1938 and Samuelson, 1947.
7. This social welfare function goes back to Mill, 1921 and is sometimes referred to as classical or additive utilitarianism.
8. This is sometimes referred to as the *Rawlsian* social welfare function after Rawls, 1971.
9. This is the classic case of Edgeworth, 1881.
10. See Mirrlees, 1971, Roberts, 1984 and Tuomala, 1990. Taking occupational choice into account, it is even more limited. In fact, rather perverse redistributive results can be achieved in this case, as has been shown in Boadway, Marchand and Pestieau, 1990.
11. For a detailed discussion of the role of the corporate tax, see Boadway, Bruce and Mintz, 1987. There are, of course, those who see the corporate tax as a vehicle for taxing the rich and would like to see more revenues extracted from the corporate sector. They would do

this through the corporate tax itself as well as through such things as minimum taxes levied on corporations. It is not clear that thinking of the corporate tax as a redistributive tax makes much sense. For one thing, it would be a very blunt instrument for the purpose. For another, it has become widely accepted that, given the degree of openness of international capital markets and the ease with which corporate capital can be reallocated among countries, much of the incidence of the corporate tax falls back on non-capital factors of production, especially labour.

12. In fact, many of these are now delivered through credits rather than deductions, though credits can be structured to have similar effects to deductions when combined with rate structure changes. We would argue that, where it is appropriate to make adjustments in the tax through the base, this should be done by deductions (exemptions) rather than credits.
13. Most incidence studies have shown that the tax structure as a whole is not too different from proportionality. See the recent survey of the literature in Whalley, 1984.
14. The choice of a destination or an origin basis would be irrelevant for this equivalence.
15. Technically, a separable utility function may be written

$$U = U[f(X_1, X_2, \dots, X_n), L]$$

where X_i is the consumption of good i and L is leisure. The implication of this function is that relative demand for the goods is independent of the amount of leisure taken.

16. On theoretical grounds, the only source of market failure that has been identified as a potential argument for government intervention is the existence of search externalities (Diamond, 1981). This could hardly form a justification for the structure of UI as we see it today.
17. See Browning, 1975.
18. See Gordon and Varian, 1988.
19. The result seems first to have been discovered by Nichols and Zeckhauser, 1982. A cursory summary of this literature is found in Blackorby, 1990. The application to health services is in Blackorby and Donaldson, 1988; and that to education and pensions is in Boadway and Marchand, 1990. Guesnerie and Roberts, 1987 had applied a similar analysis to the case of minimum wages as a way of inducing more leisure. The analysis has been extended to unemployment insurance by Marceau and Boadway, 1991.
20. I have tried to deal more fully with these issues in Boadway, 1991.
21. This argument is developed in Boadway, 1989.
22. By necessities, we mean commodities which comprise a lower proportion of expenditures of a person's income as income rises. Conversely, luxuries are a higher proportion of expenditures as income rises.
23. This may be only partly true since the cost of fuel may be above its marginal cost due to taxes.
24. That is, they are free of charge at the margin. Medical plans may require premiums to be paid, but these do not vary with treatment.

25. It is also the case that persons in remote areas are better off than if they had to move to populated areas. However, that is not the issue here.
26. See Economic Council of Canada, 1982.
27. That is not altogether true. Being elderly is a part of the life cycle that everyone has a possibility of encountering. Some persons fare better than others in terms of well-being when elderly.
28. A good survey of the literature may be found in Tuomala, 1990.

REFERENCES

- Arrow, K. J. 1951. *Social Choice and Individual Values*. New Haven: Yale University Press.
- Bergson, A. 1938. "A Reformulation of Certain Aspects of Welfare Economics." *Quarterly Journal of Economics* LII, pp. 310–34.
- Blackorby, C. 1990. "Economic Policy in a Second Best Environment." *Canadian Journal of Economics* XXIII, pp. 748–71.
- Blackorby, C. and D. Donaldson. 1988. "Cash versus Kind, Self Selection and Efficient Transfers." *American Economic Review* 78, pp. 691–700.
- . 1990. "The Case Against the Use of the Sum of Compensating Variations in Cost-Benefit Analysis." *Canadian Journal of Economics* XXIII, pp. 471–94.
- Boadway, R. 1989. "Federal-Provincial Fiscal Relations in the Wake of Deficit Reduction." In *Canada: The State of the Federation 1989*. Edited by R. Watts and D. Brown. Kingston: Institute of Intergovernmental Relations, pp. 107–35.
- . 1991. "Constitutional Assignment in a Federation: An Economic Perspective." In *Canada's Constitutional Options*. Edited by R. Watts and D. Brown. Kingston: Institute of Intergovernmental Relations, forthcoming.
- Boadway, R. and N. Bruce. 1984. *Welfare Economics*. Oxford: Basil Blackwell.

- Boadway, R., N. Bruce and J. Mintz. 1987. *Taxes on Capital Income in Canada*. Toronto: Canadian Tax Foundation.
- Boadway, R. and M. Marchand. 1990. "The Use of Public Expenditures for Distributive Purposes." Queen's University Discussion Paper No. 796.
- Boadway, R., M. Marchand and P. Pestieau. 1990. "Optimal Linear Income Taxation in Models with Occupational Choice." *Journal of Public Economics*, forthcoming.
- Browning, E. K. 1975. "Why the Social Insurance Budget is Too Large in a Democratic Society." *Economic Inquiry* 13, pp. 373-88.
- Diamond, P. 1981. "Mobility Costs, Frictional Unemployment, and Efficiency." *Journal of Political Economy* 89, pp. 798-812.
- Economic Council of Canada. 1982. *Financing Confederation*. Ottawa: Supply and Services Canada.
- Edgeworth, F. Y. 1881. *Mathematical Psychics*. London: Routledge and Kegan Paul.
- Fleming, M. 1952. "A Cardinal Concept of Welfare." *Quarterly Journal of Economics* 66, pp. 366-84.
- Gordon, R. H. and H. R. Varian. 1988. "Intergenerational Risk Sharing." *Journal of Public Economics* 37, pp. 185-202.
- Guesnerie, R. and K. Roberts. 1987. "Minimum Wage Legislation as a Second Best Policy." *European Economic Review* 31, pp. 490-98.
- Harberger, A. C. 1971. "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay." *Journal of Economic Literature* IX, pp. 785-97.
- Kaldor, N. 1939. "Welfare Propositions and Interpersonal Comparisons of Utility." *Economic Journal* XLIX, pp. 549-52.

- Marceau, N. and R. Boadway. 1991. "The Redistributive Role of Minimum Wage Legislation and Unemployment Insurance." Queen's University Discussion Paper.
- Mill, J. S. 1921. *Principles of Political Economy*. London: Longman's, Green.
- Mirrlees, J. A. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies* 38, pp. 175–208.
- Musgrave, R. A. 1959. *The Theory of Public Finance*. New York: McGraw Hill.
- Nichols, A. L. and R. J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *American Economic Review* 72, pp. 372–77.
- Rawls, J. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Roberts, K. 1984. "The Theoretical Limits to Redistribution." *Review of Economic Studies* 51, pp. 177–95.
- The Royal Commission on Taxation. 1966. *Report*. Ottawa: The Queen's Printer. (The Carter Report.)
- Sadka, E. 1976. "On Progressive Income Taxation." *American Economic Review* 66, pp. 931–35.
- Samuelson, P. A. 1947. *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press.
- Scitovsky, T. 1941. "A Note on Welfare Propositions in Economics." *Review of Economic Studies* 9, pp. 77–88.
- Sen, A. K. 1970. "The Impossibility of a Paretian Liberal." *Journal of Political Economy* 78, pp. 152–57.
- . 1973. *On Economic Inequality*. Oxford: Clarendon Press.

—. 1977. "Social Choice Theory: A Re-examination." *Econometrica* 45, pp. 53–89.

Stiglitz, J. E. 1987. "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." In *Handbook of Public Economics*, Edited by A. J. Auerbach and M. Feldstein. Amsterdam: North-Holland. Vol. 2, Chap. 15.

Tuomala M. 1990. *Optimal Income Tax and Redistribution*. Oxford: Clarendon Press.

Whalley, J. 1984. "Regression or Progression: The Taxing Question of Incidence Analysis." *Canadian Journal of Economics* XVII, pp. 654–82.

TRANSPORTATION AND ECONOMIC DEVELOPMENT: A SURVEY OF THE LITERATURE

David W. Slater

July 1991

OVERVIEW

This paper is both a survey of the literature on transportation and economic development, and an essay that addresses many of the questions that are of interest to the Royal Commission on National Passenger Transportation. This paper is referred to in the text as "the survey," and it deals with the following:

- transportation and economic development in general, and as it concerns passenger transportation in particular;
- transportation and regional economic development in general, and as it concerns passenger transportation in particular; and
- transportation and tourism.

After introductory and methodology comments, these areas are developed in Parts A, B and C. An extensive bibliography is included, and is referred to throughout the paper.

THE CHALLENGE

Economic development is defined both narrowly and broadly. Narrowly, the concern is with the growth of real Gross Domestic Product (GDP), real GDP per capita, and real GDP per worker. Broadly, the following indicators are considered: GDP and some distributive measures; GDP and non-market economic activities; GDP adjusted for changes in resource stocks, indicators of social well-being and environmental quality.

The overarching set of questions are:

Has transportation caused economic development? Have identifiable transportation elements, individually or in concert with other elements, caused identifiable changes in Canadian measures of economic development? More specifically, have passenger transportation changes caused economic development?

As Maddison (1989) has shown, Canada, along with other Western industrialized economies, has gone through four phases of economic growth during the 20th century. These are:

- fairly rapid economic development before World War I;
- slow growth, stagnation and depression from the end of World War I through 1950 (for Canada and the United States, the results of the Great Depression were reversed, and considerable increases in economic activity took place during World War II);
- a so-called “Golden Era” of exceptionally rapid economic development from 1950 through the early 1970s (with considerable convergence of national productivities toward the leader, the United States); and
- slow growth and higher inflation between the mid-1970s and the end of the 1980s.

Was transportation a contributing cause of these phases of economic development? Is there anything in a recent argument that the slowdown in investment in infrastructure (particularly in transportation) since the early 1970s has been a significant factor in the recent general economic slowdown?

Also, is there anything in the contention that deficiencies in infrastructure, including transportation facilities, are a major drag on the growth prospects of the 1990s? (Munnell (ed.), 1990.)

These questions may be asked on a less macro basis. For example, have broad *sub-groups* of transportation investment, operation and regulation caused identifiable outcomes for important *segments* or *regions* of economic development? Do these sub-groups of changes add up to important changes in both the size and structure of the Canadian economy? Some illustrative questions follow:

- Has the construction and operation of the Interstate Highway System in the United States and the equivalent system in Canada since 1950 been a major contributor to national productivity in these countries?
- Historians (Bothwell et al., 1989) have recorded that the widespread use of privately owned and operated cars in Canada occurred after World War II. Was this a factor in Canada's post-war Golden Era of economic development? (*Historical Atlas of Canada*, Vol. III, Plate 53)
- Have containers and piggy-back services substantially improved medium- to long-distance transportation of medium-weight, medium-value goods over long distances, and thus contributed to growth of markets, division of labour and productivity of important economic sectors?
- Has the development of efficient medium- to long-distance jet passenger aircraft revolutionized middle- to long-distance passenger travel? Have these changes greatly improved business travel and thus made possible the efficient management of larger national and international businesses, and thus improved productivity in many sectors of the economy? (*Historical Atlas of Canada*, Vol. III, Plates 53, 54 and 55)
- Have the United States and Canada achieved efficient highway pricing and investment as well as efficient airport pricing and investment? If not, what have been the costs of the inefficiencies? What can be done about them?
- Have the regulation and deregulation of rail and trucking services reduced the inefficiencies in these services, and thus greatly improved the productivity in some sectors of the Canadian economy?

- What are the alternatives in financing mixed private-public goods such as transportation, and in particular passenger transportation? What has been used in Canada? Are the approaches efficient and equitable?

The analysis can also be carried on at a much more micro level. Have particular projects in transportation been clear successes or failures? What are the criteria for decision making and evaluation? Are they appropriate or inappropriate? Are there major gaps in the available information, so that decision making and evaluation of transportation projects are clouded with uncertainties? What are the precise linkages of transportation projects or decisions and specific economic outcomes? How has the transportation system, its use and regulation adapted to changes in technology, demands for services, and other economic circumstances?

THE LITERATURE THAT PROVIDES EVIDENCE ON TRANSPORTATION AND ECONOMIC DEVELOPMENT

A wide variety of materials should be considered, some of them cursorily. They include:

- theories of international and inter-regional trade, and of location of firms and industries;
- theories of economic development, particularly regarding capital, knowledge and the extent of markets;
- theories of public goods and public choice, and impressionistic application to transportation; issues of externalities;
- comparative economic history:
 - general literary economic history; metropolis literature;
 - quantitative economic history, particularly that which treats economic development, transportation, trade and movements of people;
 - history and analyses of shifting structures and location of economic activities; and
 - transportation history;
- economic geography, particularly that which deals with changing location and interrelationships of economic activities and people;

- empirical analyses of general macroeconomic development;
- empirical analyses of models of transportation and economic development: macro, segments, regions, also demand for and cost of analyses of transportation services, particularly of passenger services;
- decision making on capital projects, particularly on infrastructure, and, within this category, especially on transportation;
- policy literature on economic development: general, regional, and regional disparities;
- policy literature on transportation; and
- regulation, deregulation, nationalization, privatization, subsidization and taxation of transportation.

THEMES IN THE LITERATURE

One central theme is that investing in transportation has contributed to economic development. The improvements to railroads and ships after 1870, trucking in the 20th century, and the Interstate Highway System after World War II illustrate this theme.

Transportation elements are one set out of many sets of contributors to economic development. Others include saving and the accumulation of capital; improvements in knowledge and innovations; increases in the size of the labour force and improvement in its skills; and changes in trading environments.

The contributions of transportation to economic development are difficult to identify, for several reasons. To be precise about the causality, and the strength and efficiency of the contributions, is a demanding task, fraught with uncertainty.

When many factors are causes of economic development, individually and in concert, untangling the contributors is difficult. This is so whether the elements be the size and skills of the labour force, the size and structure of the stock of capital, the state of knowledge, or the trading environment, to name a few.

Transportation decisions are often made on a case-by-case basis, yet their effects often accrue on a network basis. Thus, the consequences of transportation decisions may be quite different (sometimes better and sometimes worse) than were envisioned in the initial decision.

Transportation elements are sometimes causes of economic developments, and sometimes consequences of them. The developments of cars, highways and air services have caused changes in passenger transportation. However, the increases in wealth and income of people, from whatever sources, have caused increased travel. Analyses that can cope with and measure two-way causality are notoriously difficult.

The measurement of costs and benefits of activities carried out through market economies can be made within tolerable margins of error. But the measurement of costs and benefits of government activities is more complex and subjective. Even more treacherous is the measurement of costs and benefits of saving time, reducing noise, increasing safety, restructuring logistical arrangements, and protecting the environment.

Despite the difficulties of identifying and measuring the costs and benefits of transportation elements for economic development, the effort must be made. The attempt can be carried out at a highly macro level (all transportation investment and the national output of goods and services), at intermediate macro levels (blocks of transportation and segments and regions of a nation), and at micro levels (the evaluation of particular transportation projects). Indeed, the literature to be surveyed covers all these approaches.

GEOGRAPHIC DISTRIBUTION AND MOVEMENT OF PEOPLE

Economic development generates geographical distributions of activities and people that are uneven and specialized. Geographical distribution is partly a reflection of transportation services, but is largely determined by other factors. Except for resource industries, strong tendencies have emerged throughout the world in this century for activities to be concentrated in metropolitan areas. (See Careless, 1979; Kerr, Holdsworth and Matthews, 1990; *Historical Atlas of Canada*, Vol. III; McCann, 1987.) As cities tend to differ in the types of services and commodities they produce, inter-regional movement of goods and people is necessary in a developed economy.

Transportation of people is an integral part of the growth and operation of developed economies. Much of the movement is intra- rather than inter-regional: going to and from work, shopping, attending school, obtaining health care, enjoying day-to-day recreation. Private cars and small trucks, taxis and public urban transportation are the main sources of local transportation.

Inter-regional movements of people are the main focus of the Royal Commission. As shown in its Interim Report, private recreational and vacation travel are the largest portion. Private cars are the dominant mode for short-to-medium distances, and air services for medium and longer distances. Buses and trains are now the minority modes for such travel. Inter-regional business travel, typically by air, is also substantial.

For movements of people, cars, trucks and buses are joint users of the road systems. Joint use of airports for goods and people movement is also important.

Developments in *passenger transportation* have been factors contributing to economic development. These include, for example, the increase in ownership and use of passenger automobiles in the 20th century, and post-World War II development of large, jet-engined passenger aircraft. However, just as for goods, and for the same reasons, the precise links between passenger transport and economic development are more difficult to determine. Transportation causing economic development — and the obverse, economic development causing transportation — are important characteristics of movements of people.

CONTENTIOUS ANALYTICAL AND POLICY VIEWS

Inevitably, there are considerable imprecisions about policy regarding transportation, reflecting:

- difficulties in determining cause-and-effect relationships between elements of transportation and economic development;
- differences of view about transportation services as public goods;
- difficulties in measuring costs and benefits, as noted already;
- the domain being considered — country, state or province, region, or municipality;

- different views about the feasibility of imposing and collecting charges for use of public facilities;
- differences of market conditions;
- differences in values or objectives of economic and social policy; and
- differences of view about the efficiency of government.

From even a preliminary review of the literature, some contentious analytical and policy views emerge:

- the United States and Canada have some transportation facilities and services that fall short of current technology, demands, and cost opportunities;
- even when facilities are appropriate, they are used inefficiently;
- counting of the benefits and costs of transportation is inappropriate;
- the nation has too much, or too little infrastructure, particularly transportation infrastructure;
- decision-making models for transportation are grossly wrong;
- efficient transportation services are impeded by massive monopoly practices, “feather-bedding,” subsidization, regulation, and bureaucratization; and
- efficiency considerations in transportation are overwhelmed by considerations of entrenched special interests, distributional pressures and political values (see the debates between Winston, 1990 and Altshuler, 1990 in Munnell (ed.), 1990).

THE SURVEY

PART A — TRANSPORTATION AND ECONOMIC DEVELOPMENT

THE STATISTICAL RECORD OF ECONOMIC DEVELOPMENT

Maddison provides the most thorough and up-to-date analysis in his study, *The World Economy in the 20th Century* (1989). What follows are a few extracts from that book.

Table 1 is a summary of the aggregate of the 16 Organisation for Economic Co-operation and Development (OECD) countries and some components, 1900–87. It shows that per capita real incomes among the OECD countries were over five times as large late in the 20th century as at its beginning. It also shows that about two thirds of the increase in real product was from increases in output per capita.

Table 1
SUMMARY OF THE AGGREGATE OF THE 16 OECD COUNTRIES AND SOME COMPONENTS — 1900–1987

(a) GDP in billion "international" dollars at 1980 prices	1900	603.1
	1987	7,759.3
(b) Population (million persons at mid-year)	1900	310.0
	1987	700.7
(c) Per capita GDP in "international" dollars at 1980 prices	1900	1,946.0
	1987	11,073.0
(d) Rate of growth of GDP (annual average compound rate) 1900–87		3.0
(e) Rate of growth of population (annual average compound rate)		
1900–50		1.3
1950–87		0.5
1900–87		0.9
(f) Rate of growth of per capita GDP (annual average compound rate)		
1900–50		1.1
1950–87		3.3
1900–87		2.0

Table 2 is an extract from Maddison. It shows that trends in Canadian GDP growth were well above the OECD average in the 20th century, and more so in the first half than in the second half. A major factor in this growth has been the more rapid growth in Canada’s population than the average of OECD countries, both in the first half and in the second half of the century. Canadian productivity growth has not been outstanding, being a little above the average of the OECD countries during the first half of the century, and a bit below during the period from 1950 to 1987.

Table 2

INDIVIDUAL COUNTRY GROWTH PERFORMANCE, 1900-1987

(ANNUAL AVERAGE COMPOUND GROWTH RATES)

	GDP	Popu- lation	GDP per capita	GDP	Popu- lation	GDP per capita	GDP	Popu- lation	GDP per capita
	1900-87			1900-50			1950-87		
Australia	3.1	1.7	1.4	2.4	1.6	0.8	4.0	1.9	2.1
Austria	2.2	0.3	1.9	0.8	0.3	0.5	4.2	0.2	3.9
Belgium	2.1	0.4	1.6	1.3	0.5	0.8	3.2	0.4	2.8
Canada	4.1	1.8	2.3	3.9	1.9	2.0	4.4	1.7	2.0
Denmark	2.8	0.8	2.0	2.7	1.0	1.6	3.1	0.5	2.6
Finland	3.3	0.7	2.6	2.7	0.8	1.9	4.1	0.6	3.6
France	2.4	0.4	2.1	1.3	0.1	1.2	4.0	0.8	3.2
Germany	2.8	0.7	2.2	1.7	0.8	1.0	4.4	0.5	3.8
Italy	2.8	0.6	2.2	1.8	0.7	1.1	4.3	0.5	3.7
Japan	4.3	1.2	3.1	2.3	1.3	1.0	7.1	1.0	6.0
Netherlands	2.9	1.2	1.7	2.4	1.4	1.0	3.6	1.0	2.6
Norway	3.4	0.7	2.6	2.9	0.8	2.1	4.0	0.7	3.4
Sweden	2.8	0.6	2.3	2.6	0.6	2.0	3.1	0.5	2.7
Switzerland	2.8	0.8	2.0	2.6	0.7	1.9	3.2	0.9	2.2
United Kingdom	1.8	0.4	1.4	1.3	0.5	0.8	2.5	0.3	2.2
United States	3.2	1.3	1.8	3.1	1.4	1.7	3.2	1.3	1.9
OECD average	2.9	0.9	2.1	2.2	0.9	1.3	3.9	0.8	3.0

Source: Maddison (1989), Table 1.2.

Table 3 is an extract from Maddison. For the OECD countries, it shows the enormous shift in the structure of employment and output. These data remind us of the enormous exodus from employment in agriculture to both industry and services, and more recently from industry to services. The recent shift of output from industry to services has been less than that of employment, reflecting the higher growth of productivity in industry than in services.

Table 4 is a reproduction from Maddison. It shows the exceptional rate of productivity growth among the OECD countries in what Maddison calls the "Golden Era," between 1950 and 1973. Average growth rates of productivity since then have been a little above those during the first half of the century. The data also show that Canadian growth of productivity was well above the OECD average during the first half of the century, and has been below the average since 1950.

Table 3
LONG TERM CHANGES IN STRUCTURE OF EMPLOYMENT AND OUTPUT
(OECD AVERAGE)

	Employment			Value added		
	Agriculture	Industry	Services	Agriculture	Industry	Services
1900	38	31	31	28	31	41
1950	25	36	39	15	41	44
1980	7	34	59	4	37	59
1987	6	30	64	4	36	60

Source: Maddison (1989), Table 1.4.

Table 4
OECD PRODUCTIVITY GROWTH (GDP PER MAN HOUR), 1900–86, FOR THE 16 OECD COUNTRIES,
INCLUDING CANADA

	1900–13	1913–50	1950–73	1973–86
Australia	1.1	1.6	2.7	1.8
Austria	1.5	0.9	5.9	2.8
Belgium	0.9	1.4	4.4	1.7
Canada	3.5	2.4	2.9	1.5
Denmark	2.2	1.6	4.1	1.5
Finland	2.1	2.3	5.2	2.5
France	1.6	2.2	5.0	3.4
Germany	1.5	1.0	6.0	3.0
Italy	2.4	1.7	5.5	2.1
Japan	2.3	1.7	7.6	3.1
Netherlands	1.1	1.7	4.3	1.8
Norway	2.1	2.5	4.3	3.3
Sweden	1.6	2.8	4.4	1.6
Switzerland	1.6	2.7	3.3	1.6
United Kingdom	0.9	1.6	3.2	2.5
United States	1.7	2.4	2.4	1.2
OECD average	1.8	1.9	4.5	2.2
USSR			3.6	1.2

Source: Maddison (1989), Table 7.2

Table 5 is an extract from Maddison. These data show the general trend of convergence of levels of national productivity toward the leader, the United States. There are explainable exceptions, such as the relative decline of Australia and the United Kingdom during the first half of the century, and

the setbacks for many European countries and Japan due to World War II. These data also show that the level of Canadian productivity has been second only to that of the United States throughout most of the 20th century.

Table 5
COMPARATIVE LEVELS OF OECD PRODUCTIVITY (GDP PER MAN HOUR) 1900-86
(U.S. GDP PER MAN HOUR = 100)

	1900	1913	1950	1973	1986
Australia	94	87	64	68	73
Austria	47	46	26	57	70
Belgium	63	58	40	62	90
Canada	61	76	76	85	89
Denmark	52	56	42	61	63
Finland	30	31	30	55	65
France	41	41	38	67	89
Germany	49	48	29	64	79
Italy	39	43	33	66	74
Japan	16	17	13	40	51
Netherlands	72	67	51	78	84
Norway	39	41	42	64	84
Sweden	41	41	46	72	76
Switzerland	53	53	58	70	73
United Kingdom	82	74	54	64	75
United States	100	100	100	100	100
15 country average (excluding United States)	52	52	43	65	76
USSR			28	36	36

Source: Maddison (1989), Table 7.3.

THEORIES OF ECONOMIC DEVELOPMENT

This paper is concerned with developed economies. Much interest existed in the 1950s and 1960s in explaining the variations in economic development of particular countries over time, and differences in economic development among nations. While there were differences in both the fundamentals and the nuances of these explanations, some common ground is found, particularly regarding the *proximate determinants* of economic development. Many of the examinations proximately explained *levels* of potential (sometimes actual) economic development by treating GDP as a function of the quantities

of land, labour, capital, quality of labour, and productivity (more recently, total factor productivity). (See Abramovitz, 1952; Hirschman, 1958; Kindleberger, 1965; Lewis, 1955; Hood and Scott, 1957; and Solow, 1962.)

Derived from this starting point, the *growth* in output was taken proximately to be determined by the *growth* of land in use, the *growth* of the quantities of labour and capital, the *improvement* in the quality of labour, and the *increase* in (total factor) productivity. Attempts to explain changes over time in economic development of a nation or group of nations and comparisons of development among countries were made using these proximate determinants.

For example, for the 16 OECD countries, Maddison bases his proximate explanation of the slower growth between 1973 and 1984, compared with 1950 to 1973, largely on the slowdown in the growth of total factor productivity, and secondarily on the slowdown in capital formation. (See Table 6.)

Table 6
COMPARATIVE OUTPUT, INPUTS AND PRODUCTIVITY PERFORMANCE, OECD COUNTRIES
(ANNUAL AVERAGE COMPOUND GROWTH RATES)

Period	GDP	Cropped land area	Quantity of labour input	Labour quality improvement due to education	Capital stock	Total factor productivity
1950–73	5.40	-0.40	0.54	0.40	4.72	3.35
1973–84	2.26	-0.11	-0.39	0.48	3.22	1.65

Note: Labour is in hours; the figure for labour quality improvement due to education assumes a 0.5 percent proportionate gain in labour quality from a 1 percent increase in educational attainment of the population of working age. Weights used for OECD countries to combine inputs to determine total factor productivity were labour 0.67; capital 0.30; and land 0.03. Weights are proportionate to average shares of factors in total income.

These figures have to be interpreted with care because of annual differences in utilization of inputs. The slower rates of growth and the reduced inputs used after 1973 are partly due to the greater average degrees of economic slack that followed during the years after 1973.

In the literature on economic development, these proximate determinants raise many questions. What, for example, accounts for and affects “total factor productivity,” which rather mysteriously dominates the explanations of growth? How does growth in knowledge generate increased productivity? How much of the transformation is through the embodiment of new knowledge in the stock of capital? What accounts for the continued, though variable, efforts of saving and investment? To what extent is the availability of extensive supplies of labour for small increases in real wages a factor in the differences in economic development? To what degree has the extension of markets been a major factor in explaining variations in economic development?

Many years ago, Kindleberger appraised the situation in words that still apply:

We have suggested that there is no agreement on how economic development proceeds and have implied that this is because the process is not simple. There are many variables involved, and there is a wide range of substitutability among ingredients — land, capital, the quality and quantity of labor, and technology can substitute for one another, above certain minima, although there are at the same time certain complementary relationships among them. The will to economize and organization are probably the only indispensable ingredients. For the rest, none are necessary, and none sufficient.

The writers who place greater emphasis on one ingredient or another . . . all of them have useful insights. Even the rather rigid stage theories can illuminate the development process, just as in human growth one can propound a wide variety of sets of stages. . . . But the search for a single theory of growth, or a dominant variable, or the key to development is surely too simplistic. It may be less courageous to be eclectic, but it is also more reasonable. In a complex process with many variables and wide ranges of substitution, it is foolhardy to be a true believer in one causal pattern.

Transformation of resources among sectors is a requirement of growth and development in an economy of more than one output. It is called for by the fact that consumption of any one good encounters diminishing returns after a time. As income increases, old wants take smaller percentages of income, and new wants arise. In consequence, resources must be transferred to new occupations.

The major transformation is from agriculture to manufacturing and services. Productivity is likely to differ between sectors, at any level of income, and to change at different rates. Redundant labor with no (or even negative) marginal productivity is thought to exist as disguised unemployment in overpopulated, underdeveloped countries. If demand increases in the industrial sector, the transfer of this labor into industry can hold down wages, maintain profits, stimulate industrial investment, in a particular model of development "with unlimited supplies of labor." *This model also has historical support and relevance to the recent growth of countries in Western Europe.*¹

One question that has intrigued economists and others in recent years has been the cause of the slowdown in economic growth throughout the Western industrialized countries since 1973. Even allowing for the *sui generis* growth experience of the "Golden Era" between 1950 and 1973, the subsequent decade and a half or so of growth and inflation has been disappointing. The most obvious explanation has been the decrease in investment and saving efforts.

The proportion of output saved and invested was certainly smaller after 1973 than it was before, in many OECD countries (Boltho, 1988; Dean et al., 1989; and Maddison, 1989). The consensus of respected analysts, however, is that declines in investment and saving efforts are not sufficient explanations of the overall slowdowns in the observed economic growth trends (Slater et al., 1991 paper for Investment Canada). It is noteworthy that the principal cause among OECD countries of the decline in savings efforts was not private saving, but decreased public or governmental saving in many countries.

THEORY OF INTERNATIONAL AND INTER-REGIONAL TRADE, INDUSTRIAL POLICY AND TRANSPORTATION

Neo-classical Explanations of Trade and Development

Harris's (1985) study for the Macdonald Commission is a useful basis for the limited consideration that must be given to these topics in this survey.

Harris reviews the neo-classical theory of comparative advantage and its basis on relative factor proportions. His central proposition is that countries (regions) would have comparative advantages in those goods and services

that made relatively intense use of their relatively plentiful factors of production. Countries (regions) would and should export those goods and services in which they had comparative advantages. Thus Canada would be expected to be an exporter of wheat, forest products and minerals, and to be an importer of tropical foodstuffs. Transportation costs have often been major factors in determining the degree of specialization. Thus transportation facilities and costs are a significant determinant of trade and development for primary products.

In reviewing neo-classical explanations of trade and development, Harris also acknowledges that some nations (regions) have technological advantages over others for extended periods of time, and that export specialization may reflect these advantages too.

For most of the post World War II period, the United States has been considered the overall technological leader, with the other developed countries engaging in a game of catch-up. This notion has been central to the economic development analyses of Maddison (1989) and Boltho (1988), which emphasize convergence of productivity rates of various nations toward that of the United States. One expression of the technology spread is in the product-cycle literature following Vernon (1966 and 1969).

The main theme of Harris, however, is that although the neo-classical theory of international and inter-regional trade is satisfactory for specialization in primary products, it requires major modification or additions for trade in manufactures. Harris recalls that many trade studies during the last few decades have shown that the neo-classical theory does not explain world trends in trade in manufactures.

Harris's New Paradigm

For this, according to Harris, a new paradigm is required. It should explicitly incorporate technology, highly skilled labour, and market structures that are oligopolistic and monopolistic. It should also incorporate economies of various modes of developing, controlling and exchanging information. The focus is on firms. The prize for a firm is innovation. Being first in an activity, and being well along the learning curve, offers advantages of markets and higher income. Comparative advantages can be *engineered*, that is, they

can be created by investment in research and development, skills and communications. Nations or regions with such leading activities will climb in the international league of exports, jobs and real income.

As to the location of firms that fit the new paradigm, Harris treats transportation as a relatively insignificant consideration. For most manufacturing industries, transportation is regarded as a minor cost element in production and distribution. Other factors are more important: technology, research and development, professional and managerial people, comparative labour costs and the agglomeration economies of cities.

For a country to successfully pursue Harris's new paradigm, firms have to be *plugged-in* to the international networks of research and development, applications of new technology, engineering, integration of markets, and sources of supply. Transportation and communications for these purposes may not be a large part of the costs of firms, but they are essential. *Transportation of people in the processes of production and distribution of goods and services appears to have become more important than it was in the past.*

Harris notes that:

the old perspective of looking at transport costs and distance to market as a significant determinant of location of new industry or relocation of old industry is for the most part irrelevant, given that transport costs represent a low share of total cost of most goods. Transport costs are rarely the most significant determinant of location (p. 61).²

Harris's main concern is to develop and apply this new paradigm to Canadian industrial output and trade in manufacturing. While he only examines trade in services in a cursory way, the Institute for Research on Public Policy studies of the service industries (Dobell and English, 1988) and the Fraser Institute studies of services (Grubel and Walker, 1989) suggest that the new paradigm is applicable to them also.

Harris does not differ with the leader-follower interpretation of post-World War II economic development of the Western industrialized nations (with the United States as leader), as expounded in Maddison and Boltho. Application of his paradigm, however, suggests two important qualifications. The

first is that explanations of variations among nations in their climb in the ranks of productivity performance requires careful analysis of the industrial development of the individual countries. Secondly, the position of "leader" is likely to be shared more widely in the future.

THEORIES OF PUBLIC GOODS AND PUBLIC CHOICE, AND IMPRESSIONS REGARDING APPLICATION TO TRANSPORTATION: ISSUES OF EXTERNALITIES

For decades now, most developed countries have regarded elements of transportation as public goods on one or more grounds. (See Pigou, 1918; Knight, 1924; Musgrave, 1959; Coase, 1960; Boadway, 1985; Grubel and Walker, 1989; Watson, 1988.) Grubel and Walker, for example, have noted that:

. . . social benefits of consumption exceed the amount of revenue which the producers can recover through private market sales and which are needed to cover the costs of production (for example, light-houses, roads, education, defence, police, administration of justice, fire protection, and public transportation).

Costs of collecting fees from users may be so high that incentives for private construction and maintenance of these facilities and services are inadequate to assure an efficient supply. This is often called the *free rider* problem; efficient ways do not exist to exclude non-payers from using a service, such as a highway.

There may be negative externalities, such as noise, pollution and accidents, as well as positive externalities, such as improved access, saving of time, and so on. Yet, there may not be efficient ways to collect appropriate fees from those creating the negative externalities. The public may not even recognize their individual contributions to positive or negative externalities. With respect to traffic congestion, for example, individuals may believe that their additional use of a road adds an imperceptibly small increment to the congestion; thus they may not take it into account in their decisions.

There may also be merit rather than, or as well as, pure economic value to supplying a good or service, as is the case with the contribution of education to an informed citizenry in a democracy.

Markets may also be subject to significant inherent imperfections, such as the markets for health protection insurance and unemployment. In these markets, costs of marketing and administration are high relative to the value of the insurance protection; control of moral hazard behaviour is difficult; and self-selection of risks is likely in a market arrangement.

The good or service may be a natural monopoly; without public intervention of some sort, too little of the good or service may be produced, and too high a price may be imposed. In the past, these considerations prompted public interest in postal services, railroad services, city light services, city electricity and gas services, telephone services, radio and television transmission, and airport services.

A resource may be inherently a common property, such as a fishery, an oil or gas pool, or a common ground owned and used freely by all members of a village. Unless public policy devises means of common exploitation of such common properties, the property may be degraded by individual use, to the detriment of the relevant groups of people.

It has been widely understood for a long time that, for most situations in which there is a public interest, there are many different ways in which that public interest can be exercised. Activities can be privately produced and distributed, subject to taxes and subsidies to users or producers, to regulation of service and price, and to control of supply and use. Activities can be produced by governmental agencies; and the goods or services can be distributed freely, or sold, or subjected to direct or indirect user-charges. Property rights to fish, to search for oil and gas, or to pollute can be created by governments and private groups, and they can be sold in various ways, including public auctions.

Bird (1976) provides a particularly good survey of the theory and practice of user-charges and earmarked taxes in Canada. The Gramlich paper in Munnell (ed.) (1990), "How Should Public Infrastructure Be Financed?" and the Gomez-Ibanez paper in the same volume, "What Are the Prospects for Privatizing Infrastructure? Lessons from U.S. Roads and Solid Waste," reflect U.S. experience and changing views on dealing with public and mixed public-private goods.

The interesting questions for this literature survey are:

- how common are externalities and public goods — particularly in transportation?;
- what is the basis for externalities and public goods?;
- have externalities and public goods changed in any way, particularly with respect to passenger transportation?; and
- have the feasible and efficient ways of dealing with these externalities and public goods changed?

It is not for us to provide expert answers to these questions. Specialized knowledge is needed for that work. Rather, we point to some important suggestions found in the literature.

Changing Externalities

Massive changes have taken place in the externalities and public goods aspects of transportation. The post office, the railroads, and urban transport are either no longer natural monopolies, or, if still technically natural monopolies, are subject to much more competition. The degree of monopoly power, where some still exists, is much less. Courier services, fax machines, trucking, air freight, and the private car have lessened the power of these former monopolies. The natural monopolies in communications are less secure against competition than they were formerly.³

It appears that microelectronic technology is making it possible to impose user-charges efficiently in situations where they could not be imposed in the past (such as the application of electronic surveillance and charging for urban vehicle congestion).

Successful developments in property rights, and conditions for their use, have taken place in a number of other areas: oil and gas exploration and exploitation, *stinting* rights in the east coast offshore fishery, and landing rights and terminal use at airports. Coase (1960) challenges a number of the classic cases for public goods — lighthouses, bees and pollination — with evidence of efficient development of property rights and markets. Coase shows that efficient markets develop over time but only when the market involves a relatively small number of participants.

While the literature indicates that some traditional views of externalities and public goods have become (or are becoming) less persuasive, it also suggests that some externalities and public goods are becoming more important. These include transportation.

The most prominent negative externality is the atmospheric pollution produced from the burning of hydrocarbons by cars and trucks. Noise and accidents are also of increasing concern.

Positive externalities receiving increasing attention in the literature include improvements in the time taken to produce and distribute goods, and increases in the reliability of truck transportation. Just-in-time inventory and production arrangements have received a good deal of attention in recent years, according to Hickling (1990). So have improved distribution arrangements says Quarmby (1989).

Externalities and Government Intervention

The economic case for government intervention in transportation is still compelling. Nonetheless, it has been challenged a good deal in the last three decades by concerns over the possible inefficiencies of governments. The challenge has been systematized in the theory of public choice (notably by Buchanan and Tullock, 1962. See also Watson, 1988 and Grubel and Walker, 1989).

Since transportation, historically, has been one of the major activities subject to public interest and alleged externalities, it has accumulated an enormous body of government interventions. Common observation reveals an enormous buildup of groups with vested interests in the continuation of these interventions, including businesses, politicians, bureaucrats, labour organizations, local communities and lobbyists. These groups are clearly visible whenever changes are proposed. Thus, the concern of public choice analysts with the inefficiency of government also applies, in some measure, to transportation.

With respect to infrastructure, including transportation, the battles between proponents of economic efficiency and other values and considerations have intensified recently. A particularly clear and intense example is found in the papers and comments by Winston and Altshuler in Munnell (ed.) (1990).

Winston is a well-known and respected analyst and advocate of economic efficiency in the investment and pricing of highways and airports. But, in this paper, he detects an increased concern for economic efficiency among decision makers for transportation.

Altshuler, a political scientist and experienced public official, and now Professor of Urban Policy and Planning at Harvard, challenges this view. His doubts are reminiscent of the arguments in the 1930s about the compatibility of democratic governments and effective economic organization and acceptable sharing of income. After setting out his pessimistic view of democracy in the United States, Altshuler notes:

American infrastructure policy has been far more responsive to group pressures and broad popular attitudes than to efficiency arguments, and . . . it is likely to remain so. . . . road congestion charges remain a political loser. . . . a shift in the basis of truck taxation from gross weight to axle weight . . . is plausible. . . . The more difficult question is whether trucking taxes will increase sufficiently under this scheme to bring about a major reduction in the implicit subsidy that heavy trucks have long received. Here I have severe doubts, since the general nature of the subsidy has been well understood for decades. . . .

I believe that pricing strategies to alleviate airport congestion are forthcoming. Air traffic delays are of interest to large numbers of voters as well as to commercial airlines. It seems unlikely that runway and terminal expansion, or air traffic control improvement, will be sufficient to alleviate airport congestion in the face of rapid traffic growth in the decades ahead. The great question is whether the problem will become so severe that politicians are willing to challenge the general aviation lobby. (pp. 210–11)

Similar political economy questions arise from related papers. In the public finance literature, user-charges and earmarked taxes have had much support. Gramlich's paper in Munnell (ed.) (1990) is a recent illustration. Yet user-charges and earmarked taxes are not often utilized. When they are, these systems tend to be poorly designed. Privatization of pure public goods has been recommended by many economists, but little privatization has taken place. Why? Altshuler's arguments about the values, special

interests and inertias of U.S. democracy are too extreme. Nevertheless, the limited success of economic efficiency arguments suggests that Altshuler's approach should be given careful consideration in the formulation of public policy, particularly with respect to transportation.

MACRO VIEWS ON TRANSPORTATION AND ECONOMIC DEVELOPMENT

In the literature on economic development, Kindleberger has also made some wise observations on transportation and communication. He considers increasing the size of the market important not only to the exploitation of comparative advantages of regions, but also to the pursuit of economies of scale in a world that is far from purely competitive. The linkage of markets by improvement in transportation, or by an improvement in a product that makes it lighter and more readily transported, becomes part of a development process.

"It is not surprising," says Kindleberger, "that economic development is correlated positively with transport facilities." This is illustrated by international comparisons of the density of railroad lines of nations compared with income per capita, and density of roads, compared with income per capita. Kindleberger also cites illustrations of post-1870 expansion of trade and transport, and the growth in per capita income.

Kindleberger links contributions of communications and transportation.

Too little attention is generally paid in these accounts . . . [of economic development] to the spread of communication needed to link markets. . . . Along with transport, or rather some distance in advance of capacity to transport, there must be a network of communication, which is vital to market operation.⁴

History of Transportation as a Contributor to Economic Development

It is not necessary to deal at any length or in detail with the general economic history in which transportation is given much credit as a contributor to economic development. A few reminders of that history will suffice:

- The faster growth of trade than of output throughout the developed world is well known, and it could not have taken place without improvements in transportation and communication.

- The historical economic geography of Canada in the 20th century as presented so beautifully in Volume III of the *Historical Atlas of Canada*, shows the connections between transportation and economic development in this country.
- The travel sections and advertisements in the weekend newspapers are visual evidence of the enormous growth of the tourism trade, an industry in which transportation is an essential ingredient.

DISAGGREGATED MACRO MODELS OF INFRASTRUCTURE, TRANSPORTATION AND NATIONAL ECONOMIC DEVELOPMENT

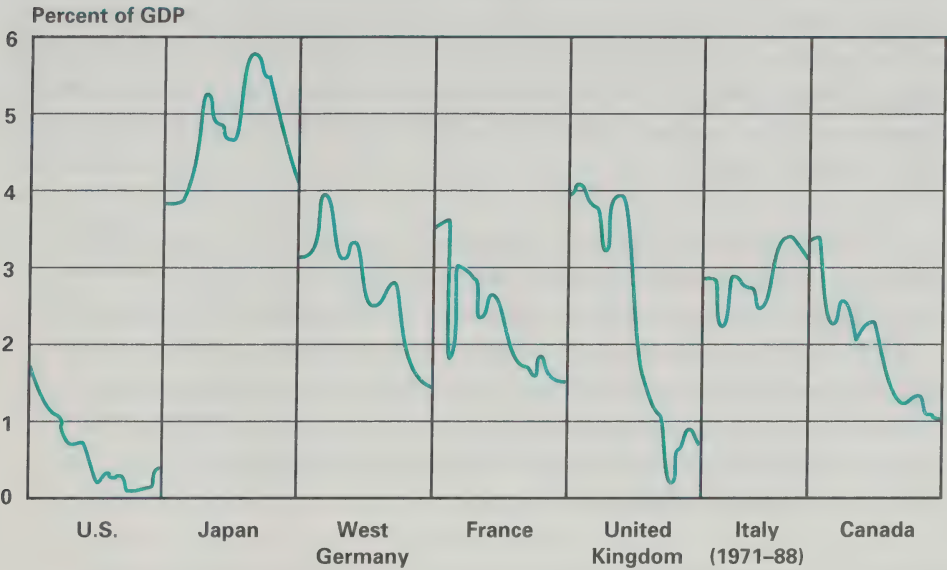
Ordinary observation⁵ has linked the deterioration of infrastructure, increased congestion, decreased investment in infrastructure and slower economic growth during the last 15 years in the United States and Canada. Infrastructure deterioration has often been attributed to decreased investments in transportation systems. This is one important cause of slower growth and smaller increases in productivity. (See Choate, 1984; Governors' Task Force, (1989; U.S. National Research Council, 1988; U.S. Congress, Joint Economic Committee, 1989; Munnell, 1990c; and Munnell (ed.), 1990a, particularly papers and comments by Aschauer, Aaron, Musgrave, Munnell, Hulten, Friedlaender, Peterson, Blinder, and Tarr.)

At the most general level of observation, public investment declined as a share of gross domestic investment in the Big Seven of the OECD members between 1967 and 1985. Chart 1 is reproduced from Aschauer (1989c). As is known from Dean (1989), gross investment as a share of GDP declined during the same period, and public investment as a share of GDP markedly decreased.

The Canadian National Income and Expenditure accounts show the decrease in government gross fixed capital formation in recent years. Table 7 summarizes the post-World War II record.

Chart 1

PUBLIC INVESTMENT AS A SHARE OF GROSS DOMESTIC INVESTMENT: 1967-1985



Source: National Accounts (OECD). Reprinted from Aschauer (1989c).

Table 7

CANADA — SUMMARY OF GROSS FIXED CAPITAL FORMATION AS A PERCENTAGE OF GNP AND GDP

Period	Government Gross Capital Formation	Business and Housing Capital Formation
As a percentage of GNP		
1950-54	3.09	18.18
1955-59	3.82	20.50
1960-64	4.21	17.20
1965-69	4.22	18.60
1970-74	3.72	18.63
1975-79	3.29	19.81
As a percentage of GDP		
1975-79	3.32	22.00
1980-84	2.88	20.04
1985-89	2.42	18.63

Note: Investment and GNP figures are calculated from Department of Finance, *Economic Review*, 1980. Investment and GDP figures are calculated from Statistics Canada, National Income and Expenditure Accounts, March 1991 (latest number).

INVENTORIES OF DEFICIENCIES OF INFRASTRUCTURE

Inventories of the amount and quality of infrastructure deficiencies have been compiled in Canada and the U.S. by central governments, states and provinces, cities, and highway authorities. (See Choate, 1984; Congressional Budget Office, 1983; U.S. Governors' Task Force Report, 1989; Joint Economic Committee, U.S. Congress, July 1989; and Federation of Canadian Municipalities, 1985.)

These inventories of deficiencies usually involve a comparison of what exists with some specified desirable standards, such as road conditions, levels of congestion and environmental standards. Most of these standards are derived from engineering criteria rather than from economic ones. They do not usually deal with the issues of specific links between infrastructure and economic development. (See Batchelder, 1979; Hickling, December 1999; David Lewis et al., 1988; National Council on Public Works Improvement, 1988.)

Recently, however, a number of macroeconomic studies exploring the quantitative linkages of infrastructure in general, and transportation in particular, to economic development have appeared. In order to consider the relationships of particular factors to potential and actual economic growth, it is necessary to disaggregate the general models of economic growth. This is equally so for education, training and retraining programs, research and development, transportation development, and so on.

The best known of these quantitative studies are by Aschauer (six articles between 1988 and 1991 are listed in the bibliography). Other prominent studies are those by Deno (1988), Eberts (1991), Fox — a detailed evaluation of a large body of empirical work of this type — (1990), and Munnell (1990b).

The Aschauer studies focus on treating private capital and public capital (or types of public capital) as separate determinants in a production function linking outputs, such as GDP, to inputs, such as quantity and quality of labour and technological change. Various formulations of the production function are used in the articles. Surprisingly large increases in productivity are attributed directly or indirectly to public capital (for example, through the stimulus of public capital to the productivity of private capital). The studies make comparisons over time, across countries, among regions, and among

sub-sectors of economies such as manufacturing. Qualitatively similar results have been generated from these studies, such as large contributions to economic growth from infrastructure, though with substantial differences in the quantitative results.

These results have met with profound scepticism from some analysts, notably by Schultze (1990), Aaron (1990) in Munnell (ed.), Musgrave (same volume) and Winston (1991).

Notes Schultze:

According to Aschauer's regression, (March 1989) a 1 percent increase in the stock of public infrastructure raised the level of output — everything else held constant — by 0.39 percent during the period from 1949 to 1985. By virtually all estimates, that increase was larger than the gain in output from a 1 percent increase in the stock of private business capital. Yet the stock of business capital (in 1987) was 3.3 times the size of the stock of public capital. (p. 63)

Says Schultze:

Those same results also imply that a one-time increase of \$10 billion in the net stock of public infrastructure would yield a permanent increase of \$7 billion in the annual level of GNP. While not a free lunch, this would be a very cheap banquet. (p. 63)

Aaron (1990) in Munnell (ed.), notes that:

David Aschauer has made one of the more fascinating and important contributions to this debate [concerning the growth slowdown]. He has called attention to the rather extraordinary disregard by economists and others of the possible role of public investment in explaining the slowdown. He has produced a series of papers in support of his contention that a sharp deceleration of public investment, especially investment in what he calls "core infrastructure," is very nearly sufficient to explain why growth slowed in the United States. . . . The paper presented at this conference continues his efforts to marshal support for this thesis. . . . Aschauer has had a valuable insight but has greatly exaggerated its quantitative importance; this paper does little to advance the thesis he propounded elsewhere. (pp. 51–52)

According to Winston (1991):

Some readers may be familiar with the work of Aschauer (1989), who estimates time series regressions that attempt to explain the impact of the nonmilitary public capital stock on the nation's productivity, and finds very powerful effects. In fact, the effects he finds are too powerful.

Consider a one-time lump-sum \$60 billion increase in infrastructure spending. An increase of this magnitude in 1985 would enable public works capital spending to regain its 1960 share of GNP. This lump-sum investment represents a 6 percent increase in the value of the infrastructure stock. Using Aschauer's elasticity estimate of .24 for the change in productivity with respect to the change in the infrastructure capital stock leads to a 1.4 percent increase in current output from the investment or a \$70 billion gain in the first year. With conservative parameters, the present value of the gain in future years would exceed \$600 billion, for a benefit-cost ratio of 10:1.

This return is implausible. Charles Schultze (1990) argues [that] Aschauer's findings simply demonstrate that the time pattern of productivity and public investment growth are similar (both rising in the 1950s and 1960s, and both falling in the 1970s and 1980s), and that this correlation generates grossly inflated estimates of the return to public infrastructure investment. (p. 126)

Munnell (1990a and b), in careful work, has produced some more modest quantitative estimates than Aschauer; she continues to attribute importance to his general theme.

While not accepting entirely the arguments of Aschauer and company, Hickling gives some weight to the notion that the record of infrastructure investment has been a factor in productivity experience. This view appears to be based in part on other indicators of accumulated under-investment in infrastructure in general, and in transportation in particular.

The Fox Paper⁶

A brief review of the Fox Paper is warranted because it is the most comprehensive and sensible evaluation of a large body of recent theoretical and empirical work.

Despite all the limitations and difficulties of the recent macro and sub-macro investigations of infrastructure and economic output, Fox concludes that some cautious conclusions may be drawn:

- Public capital investments can increase production.
- The mechanisms through which public capital affects output include:
 - complementarity between the public capital stock and private capital;
 - public capital operating as an input directly in the production process;
 - Keynesian demand effects from construction.
- Public capital investments are subject to diminishing and potentially negative returns.
- Investments in core infrastructure, and particularly transportation and roads, appear to offer the greatest productivity gains. Water and sewer and communications are also productive.
- The relative effects of marginal public investment at a particular site depend on the structure of industry where the investments are made, characteristics of other inputs available in the region, current infrastructure levels, and demand for output that would be produced in the region. . . .
- Public capital investments may crowd out private investment, but the net effect of new infrastructure still can be greater private investment and production if the complementary relationship between public and private capital in production offsets contemporaneous crowding out. (pp. 47–48)

Fox notes that:

. . . the major limitation of the research is it has little, if any, direct policy application, though it has been useful in expanding economists understanding of infrastructure's role in the economy. . . . With continued research we can continue to expect development of sound guidelines for answering these questions and targeting investments to those

locations where they are likely to be most effective. *However, the guidelines will only assist in making the best decisions and need for careful benefit/cost analysis to consider the specific merits of individual products will remain as important as ever.* One guideline already apparent is that infrastructure needs vary widely by region and country. [emphasis mine] (pp. 48 and 51)

Fox's Approach

For those interested in technical analysis, a brief review of Fox's approach will be of interest. He begins by setting out a framework that can be used for evaluating the variety of theoretical and empirical macro literature.

One part of the framework is a generalized form of an aggregate supply function which, among other things, treats private and public capital as separate inputs. Another part is an aggregate demand function that is a mixture of full-employment growth and Keynesian elements, but with explicit distinct treatment of private and public investment demand. A third part is a function generating public capital stock and investment. Along with some other secondary functions, but not with a complete simultaneous equations specification, Fox uses the interaction of the aggregate supply and aggregate demand functions as the determinant of the infrastructure-output relationships.

Fox examines a large number of empirical studies of various aspects of these relationships, which are based on time-series and cross-section data, nationally, internationally and regionally. After a careful critique of these studies, he comes to the conclusions set out above.

THE LITERATURE ON TRANSPORTATION AND ECONOMIC DEVELOPMENT

A considerable amount of literature now exists under this heading. It comprises books and articles that examine theoretical, empirical and policy considerations. Much of the literature recognizes the economic benefits of transportation, as well as the military, social and political benefits. As to economic benefits, some are conventionally counted in national income. However, many economic benefits that are important are not so measured, including time saved outside working hours; congestion reduced; safety improved; and emissions reduced.

The literature on transportation and economic development is generally about *the comparison of economic costs and benefits* of the construction and operation of various kinds of infrastructure. A project is considered to be a net contributor to economic development if its benefits, comprehensively measured and properly discounted, exceed its costs, also comprehensively measured and properly discounted. The literature considers the value of economic development of projects in such diverse fields as waterways, ports, irrigation, soil conservation, railroads, highways, and urban transportation. Much of the literature since the end of World War II has been concerned with developing countries. The World Bank has been an important centre of such studies. But the concepts and principles of measurement of infrastructure contributions to economic development (including transportation) are essentially the same for developed countries as they are for developing countries.

While the application of the cost-benefit approach to infrastructure may be more difficult, because of the absence of markets in many of the benefits, and the conditions of use, the consensus appears to be that the same concepts and principles apply, whether one is considering the contribution to economic development of a steel mill, a shopping centre or a new office building.

A consensus among economists on the appropriate concepts and measurements had appeared by the early 1960s in such works as Eckstein (1958); Krutilla and Eckstein (1958); Fromm (ed.) (1965); Mohring and Harwitz (1962); and Owen (1957). These concepts and measurements differed significantly from those generated in the engineering literature which dominated the planning and decision making of infrastructure at that time. The economic concepts considered to be relevant today (see Hickling, 1990) are essentially updated versions of the earlier economic work. While these economic criteria appear to be given more weight now than they were earlier, the literature suggests that engineering criteria still dominate planning, decision making and evaluation. Also, when the economic criteria are introduced, major errors of application occur. The alleged results are major economic inefficiencies (see Hickling, 1990).

Adler's article, "Economic Evaluation of Transport Projects," in a volume edited by Fromm (1965), is a clear and persuasive presentation of the economic principles and their applications.⁷ The tone of his presentation is established at the outset:

There is . . . no causal relation between the backwardness of the economics of transport evaluation and the fact that until a few years ago it was virtually the exclusive domain of engineers. On the contrary, this condition is to a considerable extent due to the failure of economists to interest themselves in this area even though it is one in which close cooperation between economists and engineers is especially important. As a result, some of the most common mistakes in project evaluation result from the failure to apply economic criteria correctly or at all; a few of these, such as the failure to distinguish between private and public costs and benefits and between average and marginal costs, are discussed below. . . . (p. 171)

The basic purpose of the economic evaluation of a project is to measure its economic costs and benefits in order to determine whether its net benefits are at least as great as those obtainable from other marginal investment opportunities in the particular country. There are, of course, many costs and benefits other than economic ones, such as the cultural opportunities from greater travel and the military and administrative advantages, and sometimes disadvantages, from greater mobility. These are not considered here . . . [for various reasons. . . .] (p. 173)

It is sometimes stated that the value of a project should be measured by its contribution to the growth of national income as conventionally measured. This is not inconsistent with the above formulation, but it is not a practical approach. For one, it would exclude certain benefits altogether, such as greater comfort from an improved highway, or the time saving used for more leisure, which would not be reflected in national income. More important, the national income approach is too complicated and indirect. . . . However, the national income approach is useful in focussing on costs and benefits from the point of view of the economy as a whole and not merely of the parties directly involved. In this way it helps in selecting the benefits to be included and those to be omitted and in avoiding counting the same benefit twice in different forms, such as when an improved highway reduces transport costs and increases land values. (pp. 173–74)

In order to measure economic benefits and costs and to compare them with other investment opportunities, they must be expressed

in monetary terms, which is the only practical common denominator. This presents a problem since market prices do not reflect real costs to the extent that workable competition does not prevail in major sectors of the economy. . . . [T]here are two special problems. . . . The first one arises from the fact that some transportation services by their very nature are oligopolistic or even monopolistic so that the prices charged for these services frequently have no direct relation to costs. . . . A second related problem arises from the direct and indirect subsidization of many transportation services by governments. . . . [G]asoline taxes and other charges on the beneficiaries do not cover the costs of highways (including maintenance, depreciation, interest, and administration); even where they may cover overall costs, there is usually no direct relation between specific user charges and the differing costs of the various transport services, such as those of trucks, buses and passenger cars. (pp. 174–75)

Measuring Economic Costs

Adler then turns to measuring economic costs, which he notes is substantially simpler than measuring economic benefits and can usually be limited to making adjustments in expenses. Sales and indirect taxes, licence fees and import duties should not be included in the costs, as they are not social costs. Wages ought to be measured on a social opportunity cost basis. The economic cost of capital should be on the social opportunity cost, which Adler suggests was in 1965 for developing countries at least 8 percent real and frequently more than 10 percent. The equivalent for capital-exporting developed countries would be lower at that time. Adler writes:

The problem of the appropriate interest rate can be minimized somewhat in the evaluation of many projects by expressing the results in terms of an internal rate of return on the investment, rather than in terms of [a] benefit-cost ratio. (p. 176) [The profession now considers the comparison of discounted benefits and costs to be a more satisfactory approach. Even when internal rates of return are used, they have to be ranked and compared with hurdle rates of costs of capital. Thus the internal rate of return does not permit escape from the difficulties of selecting a cost of capital.]

Adler notes that the costs of developing projects often occur sooner than the benefits. Accordingly, it is necessary to put both on a present-value basis at the same time, and the best reference point is the date when the project began.

Adler also notes that ancillary works necessary to the functioning of the main new transportation project should be included in the overall evaluation.

Measuring Economic Benefits

Adler then turns to measuring economic benefits. He indicates that this is usually much more difficult than measuring economic costs because:

- some benefits . . . such as the increased comfort and convenience . . . are difficult to express in monetary terms since there are usually no market prices for such benefits. . . .
- monetary benefits, such as reduced transport costs, benefit a great number of people over a long period of time, requiring difficult long-range forecasts. . . .
- many benefits are indirect, such as the stimulation to the economy from improved transportation; and for these benefits to materialize, investments in fields other than transport are frequently necessary. (p. 179)

He notes that the most important benefits from transport projects include:

- reduced operating expenses initially to the users of the new facility and also usually to those who continue to use the existing facilities;
- lower maintenance costs;
- fewer accidents;
- savings in time for both passengers and freight;
- increased comfort and convenience; and
- stimulation of economic development. (p. 179)

Adler emphasizes, like Mohring and Harwitz, the distinction between size of benefits and their distribution. It is likely that the benefits of transportation

developments will be widely dispersed in ways that depend on the nature of markets, government regulations and so on. This results in double-counting when measuring the size of benefits. It also gives rise to difficult substantive issues because the size of the benefits depends on their distribution. If, for example, a transportation project stimulates a large increase in use, the size of the benefits may be quite different than if usage is changed little.

The most direct benefit from a new or improved transport facility, and frequently the most important one, is the reduction of transport costs. In measuring this, Adler insists that *the proper method is a with-and-without test, rather than a before-and-after test*. Bearing in mind the dynamics of growth of usage and other factors, what would have been the profile of costs over time without the project? What will be the profile of costs over time with the project? The before-and-after test is commonly used to measure the benefits, usually resulting in gross over-estimates of the economic benefit.

The reduction in transportation costs counted should be the marginal cost of the movement rather than differences in average costs between the with-and-without situation. If, for example, a highway development diverts traffic from railroads to highways, which have lower costs, the benefit or saving is the marginal cost of saving railroad transport costs, not the difference between the average cost of movement on the highway and the railroad. Use of average costs to measure benefits may exaggerate considerably the benefits of the highway project.

Total distribution costs, not just costs of shipment, are the primary concern. If a project reduces costs of loading and unloading, storage, insurance, breakage, and so on, as well as shipping costs, these other benefits have to be factored into the evaluation of the transportation project being considered.

Also, in evaluating a transportation project, account must be taken of the new traffic that is created. Benefits in accident reduction and time savings are often crucial. Though difficult, they have to be translated into monetary terms to make a proper evaluation. Saving in inventories can be a particularly important result of a transportation project.

Adler is noncommittal about the economic development benefits of transportation. He notes that:

It is frequently assumed that all transport improvements stimulate economic development. The sad truth is that some do, some do not, and that even some of those that do may not be economically justified in the sense that there may be better investment opportunities. Each project must therefore be investigated individually and no helpful generalizations appear possible until more research may show that certain definite correlations do exist.

Before any transport improvement can be said to have stimulated economic development at all, a number of conditions must be met. The most important is showing that the economic development would not have taken place in any case even without the transport improvement. A second is that the resources used in the new development would otherwise have remained unused or used less productively. Finally, it is essential that the economic activity stimulated does not replace activity which otherwise would have taken place.

These conditions may be obvious, but it is surprising how often they are forgotten in practice. . . .

Where a transport facility does lead to increased output and the above conditions are met, the net value of this additional output is the proper measure of the economic benefit. In many situations, however, the transport facility is not the only new investment needed to achieve the increased production. This raises the problem of allocating the benefit . . . among the transport and the other investments. For this there exists no correct theoretical answer but there are at least three practical approaches. One would be not to make an allocation at all and relate the total benefits to the total investments. A second would be to annualize the other investment costs and deduct them from the benefits. A third would be to allocate the benefits in the same ratio as the transport investment has to the other needed investments. (pp. 189–90)

Comparing Costs and Benefits

Adler notes that:

Once costs and benefits have been measured in monetary terms to the full extent meaningful, the results can be put into at least three different forms:

- the rate of return on the investment;
- the benefit-cost ratio; or
- the pay-back period. . . .

While the basic ingredients — the value of the costs and benefits — are the same regardless of the final form in which they are expressed, the usefulness of the various forms is different, depending on the purpose. A short pay-back period is important where the future is unusually uncertain, where better investment opportunities are likely to arise soon, or where funds are not available on a long-term basis. . . .

Discounting benefits and costs by the opportunity cost of capital is theoretically the best way of comparing different projects. The most important disadvantage of this approach is that a particular interest rate must be chosen for discounting. . . .

This disadvantage can be minimized somewhat by expressing benefits and costs in terms of the internal rate of return on the investment, i.e. the rate which equalizes discounted costs and benefits. In this case, the opportunity cost of capital becomes important only in the marginal cases where the internal rate of return is not clearly above or below the area within which the opportunity cost of capital may be estimated to be. . . .

[T]he rate of return formula has the practical advantage that economists, financial experts and many businessmen have some concept of what an interest rate is, so that a rate of return is probably more meaningful to many audiences than a benefit-cost ratio. On balance, therefore, the internal rate of return on the investment is usually, but not invariably, the most satisfactory form in which to express benefits and costs of transportation projects in the less developed countries.
(pp. 192–94)

As noted above, the economic and financial profession now prefers cost-benefit ratio calculations over internal rates of return because they demand the ranking of possible projects. Also, the same judgement is required for the social cost of capital to decide on investment programs.

CAN GENERALIZATIONS BE MADE ABOUT OVER- AND UNDER-INVESTMENT IN TRANSPORTATION AFTER EXAMINING THE PROJECT EVALUATION LITERATURE?

Adler does not provide a general answer to this question. Individual projects, however they are decided upon, may be put through the benefit and cost evaluation. From the enumeration of the difficulties and possible mistakes, it appears that Adler believes that errors of both kinds arise. Some projects are undertaken or mooted, though a proper benefit and cost calculation would indicate that they are far from worthwhile in terms of economic value. Some projects are not undertaken, though a benefit-cost screening suggests that they would be worthwhile. It is not clear in Adler where he considers the balance of errors to lie.

Mohring and Harwitz (1962) are more daring. They write:

A limited investigation of the subject suggests that the methods presently used to measure these "benefits on existing highway use" on balance understate their values substantially. [A footnote suggests that too small a rate of interest discount is generally applied]. Those who have undertaken such studies have typically used very conservative values to estimate those user benefits which are difficult to quantify. Two groups of benefits in this category are particularly worth mentioning: (1) benefits to users of substitute facilities; and (2) the value of time saved. (pp. 18–19)

Mohring and Harwitz (1962) are also concerned with what measure of benefits would be comprehensive but not subject to *double-counting*. They write:

Presuming that all of these measurement problems had been surmounted and that a close estimate of the reduction in the time and dollar transportation costs afforded by a highway improvement had been obtained, would this estimate in fact cover *all* of its "benefits on existing highway use"? An unqualified "yes" answer seems in order . . . most of these presumed additional benefits actually involve income or substitution effects . . . or benefit transfers. Both true and internal transfers of these benefits undoubtedly take place. . . . *However, it must be remembered that a transfer represents the passing-on of a benefit and not the generation of an entirely new benefit* [emphasis mine].

Much of the same argument applies to the first of the substitution benefits enumerated above — that of a highway for other forms of transportation. The only net benefit is the associated saving in transportation costs, a benefit which is, to repeat, typically underestimated. (p. 22)

As indicated at the outset, Hickling's *Primer* (1990) appears to me as an up-to-date tackling of the decision-making and benefit-cost literature of the late 1950s and early 1960s. It deserves a more careful review in this survey, which is done below. Meanwhile, it is worth mentioning that Hickling's concepts, and its judgement about under-estimation of the benefits from transportation projects, are consistent with those of Mohring and Harwitz. Incidentally, Hickling favours benefit-cost ratios over internal rate-of-return measurements.

Though the regional material will be developed in a separate section of the survey, this is an appropriate place for a brief discussion of regional literature.

REGIONAL INFRASTRUCTURE (PARTICULARLY TRANSPORTATION) AND ECONOMIC DEVELOPMENT

The literature includes many books and articles on these subjects, applying both the macro and sub-macro methods of analysis, and cost-benefit analyses.

Providing that the regional definitions and measurements of output and inputs are correct, in principle, similar conclusions can be drawn about regional infrastructure and economic development as are made about national infrastructure and economic development. Indeed, many of the studies reviewed by Fox (1990) attempt just such analyses.

Similarly, in principle, cost-benefit analyses are fully applicable to regional infrastructure (and, in particular, transportation projects). In fact, since the outcomes depend on regional or local conditions and the application of other factors, regional or local venues are the preferred cases for cost-benefit analysis.

As Mohring and Harwitz (1962) argue, the fact that many of the benefits of transportation projects in one region are shared with other regions is not, in principle, a deterrent to the evaluation of the projects. Size of benefits is

one thing; distribution of them is another. Both authors recognize, however, that size of benefits may depend on distribution, and this has to be taken into account.

While, in principle, the measurements and evaluations of regional projects are achievable, in practice there appears to be a good deal of double-counting and window-dressing in the analysis of regional infrastructure and economic development.

Systematic review of the regional economic development literature and transportation is presented in Part B of this review. Among those reviewed are: Courchene (1981); Eberts (1991); regional aspects of Fox (1990); Friedlaender (1990); Freidman and Alonso (1964); Kraft, Meyer and Valette (1971); Munnell (1990a); Savoie (1986); and Sullivan et al. (1989).

Hickling: *A Primer on Transportation and Economic Development*

For the purposes of the Royal Commission, Hickling's *Primer* is the most useful piece of literature available.

- Hickling judiciously blends the macro and micro evaluation and decision-making approaches (which the *Primer* calls the indicative planning methodologies and the investment choice methodologies).
- While putting primary emphasis on goals of economic growth, economic welfare and improvements in living standards, it provides for incorporating regional and distributional objectives.
- It provides a sound basis for evaluating benefits, costs, and benefit-cost comparisons; and for dealing with discounting of costs and benefits.
- It points toward the dynamics and efficiency considerations that are appropriate to establishing the *base cases* against which new projects should be compared.
- The *Primer* provides a convincing comparison of engineering and economic criteria for evaluation and decision making for transportation projects. It also provides an impressive survey of current practices and common errors, without being "smart-alecky" or "preachy."
- It incorporates logistical and environmental implications of transportation projects.

- The *Primer* points toward sensible answers to the questions of over- and under-investment in transportation projects.
- It recognizes that we live in a world of risk and uncertainty, and have to decide upon and evaluate transportation projects accordingly.
- It is written in non-technical language.
- And finally, the *Primer* is educational rather than critical in tone.

Here are a few highlights from the *Primer*:

- “Growth for growth’s sake has never been the center-piece of American public policy. Nevertheless, the fact stands that growth, through acceptable means and at acceptable costs — sustainable development — is the only means available to recover and sustain ground in American living standards, and most of the increased growth can be achieved only through increased levels of productivity.” (p. 1)
- “There is wide-spread agreement that higher rates of capital investment are key to the future growth of productivity and living standards.” (p. 3)
- “In the public sector, where market forces are weak, special efforts must be made to ensure that infrastructure investment matches and enhances productivity gains in the private sector.” (p. 3)
- “In the case of transportation infrastructure . . . [u]nless these investments yield economic gains, including productivity gains, that exceed the costs of achieving them, they will make no contribution to the nation’s overall rate of economic expansion. Interregional competition, like competition generally, is a healthy thing. But when a region grows at the expense of others without generating a net contribution to the sum of all economic activity, living standards for all will stagnate and decline over the long-term.” (p. 5)
- “At a minimum, decision makers need to assure themselves that policies and programs will make a net contribution to economic growth. . . . [They] do not have to settle for only the most highly stimulative projects; a poorer locality might warrant project funding even though its proposal offers less potential for productivity growth than the proposal of a wealthier region. *But any project should offer at least a minimum net contribution to economic growth to stay in the running* [emphasis mine].” (p. 14)

- As a primary objective, Hickling recommends *growth in total economic welfare*, that is, “increased economic benefits that exceed the increased economic costs of achieving them. ‘Welfare’ is distinguished from ‘output’ in that welfare includes factors, such as time and safety, that have economic value but are not included in the normal accounting definitions of economic output and gross product.” (p. 15)
- “While productivity gains alone can often justify transportation investment, this is rarely the case with employment, income and other targets of regional redistribution.” (p. 20)
- Regarding negative environmental impacts: “The critical issue therefore is not whether such transportation investments are environmentally sustainable, but rather how to devise policies that make a portion of the benefits available to finance the mitigation of negative environmental impacts.” (p. 24)
- “How can a “minimum-required contribution” to economic growth be defined in practical terms? The operational approach . . . is through use of the rate-of-return concept commonly applied in investment planning generally.” (p. 26)
- “Two errors are quite common in recent practice. The first is a marked confusion between distributional aims and aims relating to economic growth. It is not unusual, for example, for the projected employment and related economic ripple effects of an investment to be regarded as evidence of economic growth and singled out for measurement as key program objectives A second common error is failure to conduct the kind of analysis needed to measure the prospective growth implications of policy and program options.” (p. 29)
- *Indicative Planning Methodologies*, which attempt to determine how much infrastructure spending is enough. Hickling reports on, and evaluates as a useful tool, the methods based on the infrastructure-productivity statistical analysis of Aschauer et al. (The Hickling paper is more impressed by the measurements of Aschauer et al. than I am.) *Hickling is cautious about the practical applications of such analysis*: “While the application of indicative planning methodologies can help executives identify gross under- and over-investment levels from a budgetary perspective, only the application of forward-looking [investment choice] methodologies can identify the most promising investments and distinguish strong from weak transportation program and project choices.” (pp. 34–37)

- *Investment Choice Methodologies* are the heart of the *Primer*: "Whether growth is defined in terms of productivity, gross output or economic welfare and living standards, it can only occur if more of value is put into the economy than is taken out (spent) in order to achieve it. Only by gauging transportation policies and investments in terms of their rate-of-return and net present value can decision makers discern their implications for productivity and economic growth. . . . [T]he state or local transportation analyst needs to ask whether the economy as a whole will be made better off by undertaking the project rather than not undertaking it, or by undertaking an alternative project instead." Hickling then compares the key measures of productivity and economic growth. (pp. 37–49)
- For a number of cases, Hickling compares commonly used engineering decision-rules with economic rules based on net-present-values. *They show that the economic measure of benefit commonly exceeds that arising from engineering decision rules.* Other things being equal, this evidence points to a bias toward under-investment in highway and airport projects in the United States. (pp. 43–44)
- "Sound economic decisions in investment planning necessitate that major new policies, programs and investments be approved only if they can be justified after accounting for the impact of developments and actions that lead to the most efficient use of existing facilities. Rarely is it the case that "nothing happens" to improve current systems in the absence of major investment. The 'nothing happens' and 'do-nothing' baselines of comparison for prospective new policies assume that the transportation system and related patterns of economic activity will reflect the status quo in the absence of investment. This assumption fails on three counts:
 - First, it fails to adjust demand for a program or a project's services to the no-investment case. . . .
 - The second problem inherent in the status quo baseline is that it ignores steps available to state and local transportation authorities to improve the productivity of transportation systems in lieu of major expansion [for example, by the use of congestion pricing].
 - The third intrinsic problem in a base case defined by status quo conditions is that it can inhibit a broad search for innovative policies and programs of solving problems. . . ." (pp. 63–65)

The remainder of the Hickling study is a thorough, and by now fairly orthodox primer on identifying benefits and costs and applying benefit-cost analysis. It is worth noting that the Hickling study considers that the time-saving and reliability of transportation have had (and can have much more) major beneficial effects in the organization of production and distribution, and in savings in inventories.

ECONOMIC DEVELOPMENT AND INTERCITY PASSENGER TRANSPORT

The literature on economic development and intercity passenger transportation is much scarcer than that on economic development, and on transportation in general. The Royal Commission has contracted for a study of the income and price elasticities of the demand for passenger transportation, one of the main concerns under this heading. Accordingly, the elasticities will be reported on here rather cursorily.

It is clear that there are two-way causations between economic development and intercity passenger transportation. Such developments as jet passenger aircraft and paved roads have caused developments in intercity passenger transportation. Looking at the relationships the other way, increased real household incomes have been a cause of the enormous increases in the ownership and use of private cars, in both intra-city and intercity passenger transportation.

Recent literature that addresses the subject includes Gillen and Oum (1981); Grubel and Walker (1989); *Mathematica* (ed.) (1966 and 1967); McRae, (January 1989); Morrison and Winston (1989); Oum and Tretheway (1988); Oum, Waters and Yong (1990); Palmer (1988); Salvas-Bronsard and Bastien (1984); Scarfe and Krantz (1988); Transport Canada (1982, 1988a, 1988b and 1989); VIA Rail (1989); and Watson (1988). The classic literature on consumer demand and the analysis of family budgets will be left to the Royal Commission's study of elasticities (such as the works of Houthakker and Taylor, and of Prais and Houthakker).

In the Royal Commission's Interim Report, the distinction is made between business and personal intercity passenger transportation.

Intercity Business Travel

Consider business travel first. Among the important points made in the *Interim Report* and other literature on this subject are the following:

- For business travel, a useful distinction is the one between travel that involves providing business services, and travel that involves other aspects of the production and distribution of goods and services.
- As Grubel and Walker (1989) show, business services account for about half of overall service economies. Also, these services have been a major factor in the rapid growth of the service economy in North America. These services make intense uses of the transportation of people and communications.
- Business travel places a premium on speed and reliability. For short-distance intercity business travel, the car is clearly the preferred mode. For medium-distance as well as long-distance business travel, air service is clearly the preferred mode. The ability to work while travelling is important, thus the rapid growth of working facilities on aircraft and phone service in cars is understandable. The saving of time is also important. Thus, for business travel, congestion is highly undesirable, particularly when getting to and using air services (Winston, 1991).
- Though price is a consideration for business travel, it is by no means the dominant consideration. The ability to obtain and adjust service quickly, and the considerations just noted, appear to be dominant. Thus, airlines have found it attractive to charge higher prices for business than for personal air service, as well as to provide better facilities.
- For business travel, complements to air service, such as airport hotels, hospitality services and airport car rentals have developed rapidly during the post-World War II period.
- During the post-World War II period, the big loser as a mode for business travel was rail passenger service.
- With improvements in service and reductions in costs, business communications are becoming a substitute for some business travel. Nonetheless, there are still some strong complementary relations between communications and business travel.

Personal Intercity Travel

The Royal Commission's preliminary work and a quick scan of the literature reveals the following points:

- Personal intercity travel is dominated by the use of personal cars for short- and medium-distance voyages. The personal use of air services is growing rapidly, particularly for long-distance voyages. Long-distance personal voyages are also growing rapidly. Train and bus services appear to be losing shares in passenger transportation.
- The spread of personal ownership and use of cars, first in America, then in Europe and Japan, has brought about one of the greatest social changes of modern times. Intercity personal travel is no longer the preserve of the rich. People of almost any age and economic standing can go almost anywhere they want, with privacy, comfort, convenience and economy. People will not readily give up their personal ownership and use of a car for a bus or train (or, in cities, for urban rapid transit). This applies more in North America, with its vast distances and low population densities, than it does in Europe or Japan. But the car, even in Europe and Japan, has a remarkably large share of the intercity transportation market.
- Car and air services are commonly labelled "superior goods," in the economists' sense of that term. The contention is that, other things being equal, increases in demand for these services will outpace increases in income. Bus service has been labelled an "inferior good" (Palmer, 1988) meaning increases in demand for bus services will not keep up with increases in income. Evidently, intercity common carrier buses are now used mainly by the young, the poor and the elderly. Palmer argues that rail passenger service in Canada is still a superior good.
- These income elasticities are important distinctions, for they provide some indication of the relative growth in the demand for various intercity passenger services as the incomes of Canadians change. The evidence, however, has to be analyzed carefully. Firstly, the demand for passenger service per household could decline or increase slowly in comparison with the increase in household income; yet the overall demand for the service could increase much more rapidly because of the increase in the numbers of households. Secondly, other factors affecting demand change over time. Quality of services may be altered by supply factors. The relative prices of the services may be changed (for example, the energy price

cycles of the last two decades affected modes of passenger transportation differently). Policies may change regarding the structure and system for imposing charges for the use of public facilities. The distribution of income and wealth may alter; personal transportation preferences may differ among income classes. Thirdly, one set of observations may arise from depressed economic conditions and another from boom economic conditions.

Industry analysts predict that intercity passenger transportation in North America will remain dominated by the private car and the aircraft during the next two decades (Johnson in U.S. Department of Transportation, 1989; U.S. Congress, Joint Economic Committee, 1989). With the application of new materials, such as composites and ceramics, and changes in design of engines (perhaps with the use of alternative fuels), the supply capabilities (quantity, quality and cost) of both cars and aircraft are expected to continue to improve. These expectations reinforce the conventional wisdom that cars and aircraft will dominate the next two decades of intercity passenger transportation.

However, these “status quo forecasts,” as Fuller calls them, may be challenged by radical changes in environmental regulations, fuel prices, and other considerations.

One key question is whether high-speed intercity rail services could become economical substitutes for car and air services. Others will have to answer that question for Canada. In my view, Hickling’s *Primer* should be applied to such issues, just as it should to highways and airports. The Hurley and Jones Discussion Paper (1990) indicates that substantial capital subsidies would be required to support high-speed intercity rail service in the Montreal–Ottawa–Toronto corridor, to say nothing of the Quebec–Montreal and the Toronto–Windsor links. Press reports of the Ontario–Quebec Task Force and of the Bombardier and Asea Brown Boveri proposals appear to call for substantial subsidies. Whether such subsidies would be justified to bring the social benefits into line with the social costs of such services, I cannot answer, for lack of data.

For conventional rail passenger service, Cubukgil and Soberman (1986) argue that subsidies would not be needed in the Quebec–Windsor corridor. They maintain that VIA Rail could operate in this corridor without them if

three conditions were met: (1) excessive charges by the railroads for their services could be eliminated; (2) feather-bedding practices of VIA Rail's workers could be cut way back; and (3) VIA Rail's excessive administrative overhead could be reduced.

EFFICIENT TRANSPORTATION INFRASTRUCTURE INVESTMENT AND PRICING POLICY

The main ideas of the theory of efficient transportation infrastructure policy were first worked out more than 70 years ago (See Pigou, 1918; Knight, 1924; and Mohring and Harwitz, 1962.) These ideas, however, remained more or less just academic musings for many years, even among economists. Recently, a tremendous amount of interest in the subject has arisen. One reason for this is that newer technology, particularly the application of microelectronics, appears to make efficient user-charge systems practical. Another has been new evidence (both data and quantitative analysis) that shows the inefficiencies of certain systems, particularly highways and airports. A third reason for the current interest in transportation infrastructure policy is the enormous estimate of the costs to maintain, restore and add to the infrastructure. This has given a renewed sense of urgency toward improving the efficiency of the system.

The literature, in addition to the classics noted above, includes: Bos (1985); Downs (1962); Bird (1976); Brander et al. (1988); Gillen et al. (1988); Hickling (1990); Oum et al. (1990); U.S. Department of Transportation (1989); Morrison and Winston (May 1990); Schultze (1990); Small et al. (1989); and Winston (1985 and 1991). The best survey of the literature is in Winston (1985), and the most compact recent application to highways and airports is in Winston (1991). As already noted, the most severe critique of the practicality of the economic efficiency advocates is by Altshuler, in Munnell (ed.) (1990).

Altshuler's central idea is to invest more efficiently in infrastructure (highways and airports appear to have received the most attention), and to then charge appropriate prices for its use.

Consider highways, for example. Two sets of problems arise, one concerning trucks and the other concerning cars and their equivalents. Trucks are the primary cause of the deterioration, and need for highway repair, resurfacing and reconstruction. The deterioration is primarily a function of the thickness

and quality of the pavement (road surface), and the axle-loading of the trucks. According to Small et al. and Winston (1991), current engineering standards do not demand thick enough and strong enough pavements. A program for upgrading pavements and charging trucks for their use (based on axle-loadings) could lead to substantial improvements in the long-run efficiency of the highway system.

Cars and their equivalents, not trucks or buses, are the main cause of highway congestion. To alleviate this, congestion pricing could be introduced as part of the regulations governing the use and financing of highways. This now appears to be technically feasible, and practical. Small et al. (1989) and Winston (1991) report on a number of successful experiments in congestion pricing.

Winston (1985) surveys the work on time and price elasticities in transportation, and indicates their importance in determining appropriate congestion pricing. These user- and congestion-charges would be in addition to gasoline and other fuel taxes, with overall revenue to cover the costs of building and maintaining the highway system. As a transition measure, some use of general revenue or borrowed funds would be required.

Highways

Winston (1991) sums up the highway case as follows:

Efficient highway infrastructure policy is designed to make the best use of scarce durability and capacity. Scarce durability arises because roads can only withstand a finite number of standard loadings before they need resurfacing. Efficient road wear pricing attempts to reduce loadings by forcing shifts to trucks with fewer loadings; efficient investment recommends road design that allows roads efficiently to withstand a greater number of loadings. Each policy extends road life and saves society maintenance expenses; together they reduce maintenance expenses even more and, most importantly, they minimize redistribution and thus political problems. Scarce capacity is effectively rationed by congestion pricing; such capacity only can be used by those motorists willing to pay an efficient premium for it. With efficient highway infrastructure policy in place, authorities are able to make efficient decisions about whether building new roads can be economically justified. (p. 122)

Winston's argument is set out more fully in Small et al. (1989), and congestion-pricing arguments are supported in varying levels of detail in Downs (1962); Hickling (1990); Johnson, in U.S. Department of Transportation (1989); Schultze (1990); and Small, Winston and Evans (1989). Though the economic efficiency arguments regarding highways are logically sound and supported by evidence, Altshuler regards congestion pricing in the United States as unacceptable for political reasons.

Airports

Similar efficiency arguments are being applied to airports, where the problem is mainly congestion. Winston (1991) writes:

Optimal airport pricing and investment policy could generate roughly \$11 billion (1988 dollars) in annual benefits. Travellers [would] reap \$8 billion in reduced delay and also would pay lower fares because the expansion in runway capacity called for under optimal investment combined with congestion pricing would reduce congestion to such an extent that, on average, landing fees would fall. The annualized cost of the additional runway investment is only about \$1.5 billion. Carriers benefit from the lower operating costs from reduced delay. Airports' net revenue would fall slightly, but, as we argue below, they would become financially self-sufficient. . . . Combining efficient pricing and investment would postpone the need to build expensive new airports. . . . Continued growth in air travel will eventually necessitate constructing new airports, but these decisions will be made more efficiently if we make better use of our current airport capacity. (pp. 123–24)

Gillen (1988) has published an elegant and technically persuasive paper on the application of airport pricing principles to Canadian airports. In particular, the paper presents the case for peak and off-peak pricing, for differentiation among types of user, and for combinations of marginal-cost pricing and demand elements in efficient pricing systems. Hickling (1990); Morrison (1983); Morrison and Winston (1989); and Morrison and Winston (1990) also present strong cases for efficient investment and pricing of airport use.

Altshuler (1990) is more optimistic about the applicability of economic efficiency principles to airport congestion than to highway congestion. Even

so, he believes that public concerns over noise and environmental considerations will outweigh much of the economic efficiency argument regarding investment in airports.

Bird (1976) presents a good case for more widespread application of user-charges on public facilities in Canada. Gramlich (1990) presents the current public finance view of the application of user-charges for public goods in the United States. Much of what he suggests would be suitable in Canada.

The efficiency literature also provides evidence as to whether there is over- or under-investment in such infrastructure as roads and airports. *It strongly supports the view that the investment has been excessive*, in the sense that if the capital stock were efficient, it could provide additional and more efficient services. Paradoxically, additional investment would be required in the short-run to improve the durability of roads, install congestion pricing systems, build more airport runways, and improve traffic control facilities. But, argue the advocates, less investment would be required over the long-run, and self-sustaining financial arrangements could be operated for roads and airports.

TRENDS AND PROSPECTS FOR TRANSPORTATION NEEDS

In this literature review, no attempt is made to canvas the forecasts thoroughly. The Royal Commission has other work under way on trends and prospects. Some incidental literature on these matters was encountered, however, and is reported on here.

Since many Canadian trends follow those in the United States, we begin with a review of those trends. We then make a few Canada-U.S. comparisons, and speculate a little on Canadian trends and prospects for transportation needs.

The U.S. literature upon which comments are made includes: U.S. Department of Transport (1989); U.S. Congress, Joint Economic Committee (1989); U.S. National Research Council, Transportation Research Board, *A Look Ahead, Year 2020*, and particularly an article by Lewis, Hara and Revis: "The Role of Public Infrastructure in the 21st Century"; Eno Foundation, *Transportation in America* (May 1990); and comments by Gail D. Fosler, in Munnell (ed.) (1990).

Demographic Trends

A good starting point is the section of the overview paper by Fuller, in *Moving America*, that deals with Johnson's paper, "Transportation for the Next Century." Fuller's paper concentrates on demographic trends as they will affect transportation, but on a more detailed basis than the Royal Commission did in its Interim Report. Fuller writes: "The cumulative impact of demographic change suggests the following conditions for the United States in the future:

- slower economic growth and relatively less total demand for new transport investment;
- a service-oriented economy with relatively less demand for freight movement;
- a premium on service in which transport speed, safety, and reliability all feature;
- problems in finding an adequate labour force for transportation;
- more demand for leisure travel;
- major highway transport demand in suburban areas of the West and South;
- more flexibility in work schedules; and
- a greater consumer premium on time and convenience in transport, making America an 'impatient society'." (p. 7)

Growth in transportation demands in the U.S. is expected to be slower in the future than in the last few decades, largely because of demographic factors. Some continued growth in demand is expected. The population is still increasing overall, and growth in the Southern and Western regions is rapid. Increases in income are expected to continue, and to eventually overtake increases in expenditures on transportation. Car and air travel are forecasted to be the preferred modes in transportation growth. The combination of that growth, the demands for better service and the economic burden of improved transportation services has increased the pressure for improved operating efficiency.

Based on Johnson's paper, Fuller contends that improvements in the operating efficiency of the transportation system are not only important, but feasible. He is more optimistic than Altshuler (1990). Fuller writes:

Operating efficiency means higher speed and improved service quality, particularly involving transport nodes, or hubs. Johnson suggests that four developments are needed to achieve efficiencies: (1) understandable service goals and measures of goal achievement, (2) *market-oriented strategies and pricing for solving transport problems* [emphasis mine], (3) new funding methods, and (4) information tools to enhance transport performance. There is special promise in applying computers to existing transportation systems; the opportunities for applying bold new technologies, however, are very limited. (p. 7)

Disagreements about the extent to which market-oriented approaches to transportation services should be adopted have been encountered already in this survey. In *Look Ahead, Year 2020*, Lewis et al. make strong arguments for increased use of user-fees for transportation services, but not exclusively user-fees. Fosler (1990) accepts the contention that rate-of-return and efficiency considerations, and privatization, may point toward increased efficiency in infrastructure investment and operations. In her judgement, however, "... infrastructure spending . . . will have to rise substantially; and it will have to rise in areas in which neither the economics nor the political process will favour private solutions." (p. 182)

Any long-run economic forecasts risk paying too much attention to some forces and not enough to others; or of not foreseeing changes that may become important. Fuller queries Johnson's orthodox forecasts. Will environmental and energy difficulties alter transportation demands or the way in which they have to be met? Will the application of computers to transportation be extensive and efficient? Some queries from other literature include: Will the Southern and Western United States be able to cope with water shortages that could limit their growth? Will the suburbanization and diffusion of the location of economic activities continue? Will society cope effectively with the transportation needs of the elderly and persons with disabilities? Will high-speed, guided ground transport in major intercity corridors be intensively developed in North America, and, if so, will this succeed in displacing much of car and air travel?

When comparisons are made of the current situation and recent trends in passenger transportation in Canada and the United States, several similarities and differences stand out (see *Transportation in America*,

May 1990; *Getting There: Interim Report of the Royal Commission on National Passenger Transportation*, 1991; Bothwell, Drummond and English, *Canada Since 1945*, revised edition, 1989.)

The post-World War II growth and age-structure of the two populations are similar, except that the rate of growth, relative levels of immigration and size and duration of the Baby Boom were larger in Canada.

The aging of the populations is similar, except that changes in Canada lag behind those in the United States. The enormous increase in female participation in the labour force has been similar too, though to a slightly greater degree in the United States. In both countries, use of private cars and air travel have become the preferred modes of intercity travel since 1945. Levels of car ownership are now high in both countries, but more so in the United States than in Canada. Similarly, use of air travel is greater in the United States. Due to the aging of the populations, growth of intercity travel is expected to slow in both countries; but income, demand, and technological changes suggest that some growth in intercity passenger transportation is likely in both. There is likely more potential for such growth in Canada than in the United States.

Canada does not have a sunbelt comparable with the one in the United States. It acts as a catalyst for population growth, and people travelling in this area rely on cars, small trucks and aircraft for transportation. Urban and suburban trends in Canada are largely based on car transportation, and it is difficult to envisage this changing during the next two to three decades. Trends toward a service-oriented economy, a premium on speed, safety and reliability in transportation, two- and three-worker households in which each person depends on his or her car, increases in leisure travel, and a predicted scarcity of young people — all these are qualitatively similar in Canada and the United States. The transformation of air service to hub-and-spoke patterns is also common to both countries.

Among the differences, proportionately more people and activities in Canada are located in areas with cold, harsh winters. Highway deterioration from weather is thus a bigger problem in Canada. Canadian tourist travel to the United States is proportionately larger than that of Americans to Canada (and absolutely larger, too).

SOME TENTATIVE CONCLUSIONS ON TRANSPORTATION AND ECONOMIC DEVELOPMENT, WITH SPECIAL EMPHASIS ON PASSENGER TRANSPORTATION

The conclusions to this survey are presented in two parts. The first is an overview of the evidence on transportation (in general, and passenger in particular) and economic development. The second is a recall of the more startling and outrageous recommendations for policy changes in transportation and economic development.

Overview of Evidence on Transportation and Economic Development

Both for transportation in general and for passenger transportation in particular, a vast and varied literature supports the contention that transportation has been a major contributor to economic development. Common observation provides a worthwhile body of supporting evidence, but there is much more. Historical and econometric analyses point to these conclusions. Much systematic evidence has been developed at macroeconomic levels, and for regions and sectors. Rigorous methods and useful standards have been applied to thousands of individual projects and proposals. Some reconciliation of engineering and economic methodologies has taken place. Despite the criticisms and calls for reform in policy regarding transportation and economic development, the conclusion set out above appears to be warranted.

The evidence suggests that transportation investment and maintenance will be worthwhile contributors to economic growth in developed countries in the future. However, those contributions do not seem likely to be of overwhelming importance on the scale of the railroad and steamship in the 19th century, or of the car, truck, highway and airplane in the 20th century.

The issue of *improved economic efficiency* in the investment and operation of transportation is raised in much of the literature, and many suggestions are made on how to reach that objective. These include congestion pricing or taxation; increased use and redesign of user-charge systems; and substitution of privatized for public services. More rigorous and careful application of economic benefit-cost measures are suggested, along with integration of the economic and engineering criteria for decisions on transportation. Still, the economists' zeal for *economic efficiency* continues to be ignored by many, and vigorously challenged by others.

The so-called “public good” arguments for government involvement in transportation (as owners, operators, financiers, subsidizers and regulators) appear to be less applicable and more selective than they were in the past. Natural monopoly elements, where they continue to exist, are subject to more competition. Changes in technology have made externalities that could not be charged to users in the past, now chargeable and collectible. The inefficiencies of governments weigh more heavily in public choice than before. Public concerns still arise in transportation services, and some have become more important (for example, environmental protection).

Speed and reliability of service are of increasing concern in transportation, but they appear to be largely determined by private decisions.

Contentious Analytical and Policy Views in the Literature

Early on in this survey, a number of contentious analytical and policy views were flagged. What follows is a summary of some of the evidence for these views.

First, many writers suggest that the United States and Canada have mixes of transportation facilities and services that are inappropriate because they are out of keeping with current technology, demands, and cost opportunities. As an example of this, see Friedlaender’s and Walters’ judgements (in Munnell (ed.), 1990) concerning the gross over-building of roads in many rural U.S. areas, as well as the over-building of rapid-transit systems to and from city centres.

Hickling shows that too little attention is paid to net economic benefits in decisions about the amount and kinds of transportation projects. Hickling also notes that misleading technical, local and regional criteria are widely used in decision-making models, particularly by state and local governments. Projects have been built that have little value. Projects with net economic value have not been undertaken because their measurement was not done correctly.

Small et al. and Winston provide convincing evidence that the design of highways (particularly the thickness and durability of pavements) is inadequate, and that the number of runways at airports is also inadequate.

Downs and Schultze argue convincingly that if transportation facilities are provided to the public at a zero or near-zero private marginal user-charge, then projects aimed at reducing congestion will inevitably cause more usage and more congestion.

A second startling contention in the literature is that facilities, even when they are appropriate, are used inefficiently. Schultze, Winston and Small et al. show convincingly that, for highways, there are inappropriate user-charges for trucks, and an absence of appropriate congestion charges for all types of users. Recently, other literature has supported instituting user-charges. For example, both Tobin and Blinder, in their testimony to the U.S. Congress Joint Economic Committee (1989), urged much greater application of user-charges to infrastructure. Hickling's discussion of the base case against which new projects should be compared is notable in this connection. Similar arguments were developed by Adler, in Fromm (ed.) (1965).

In the literature, a third startling and repeated contention is that the counting of benefits and costs of transportation is frequently inappropriate. Hickling argues that too little attention has been given to benefits such as time-saving, reliability, user-costs, vehicle damage, and safety, so that benefits are frequently under-estimated. Small et al. and Winston argue the same point.

Hickling argues that too little weight is given to the benefits arising from improved efficiency in the production and distribution of goods (such as just-in-time logistics of production and enormous economizing on inventories at all levels of production and distribution). This point is strongly reinforced by Quarmby (1989). Pollution Probe, Suzuki and others argue another case, that gross under-estimates have been made of the costs of environmental degradation caused by transportation practices. Fourthly, it was noted earlier in this survey that there were strong views in the literature that the United States and Canada have too little infrastructure, particularly transportation infrastructure. There were equally strong views that the bias was toward too much investment in infrastructure.

Evidence for there being *too little investment in infrastructure, and transportation in particular*, includes the following:

- the Aschauer et al. macroeconomic studies of the high productivity of infrastructure, a productivity that is alleged to be much above the returns

on other investment, and by implication above the social opportunity cost of capital;

- the widely held view, among both engineers and economists, that investment in infrastructure during the last two decades has been reduced, and that it has not been sufficient to maintain the real stock of that capital;
- the carefully measured judgement of such transportation economics experts as Mohring and Harwitz about the strong tendency to underestimate the benefits of highway projects;
- Hickling's illustrations of possible projects for which the benefits appear to clearly exceed the costs;
- Hickling's evidence on widespread under-estimation of benefits, particularly of highway and airport projects; and
- Hickling's contention that too high a hurdle rate (social opportunity cost of capital) is used to evaluate transportation projects.

Evidence for there being a bias toward *over-investment in infrastructure* includes:

- the convincing evidence of Small et al. and Winston that large inefficiencies in transportation infrastructure could be overcome by an initial investment to upgrade the systems, together with user-charges;
- the widespread tendency, noted by Adler and Hickling, to count as benefits things that are not benefits (such as employment), and to double-count benefits (transportation savings plus increments to land values); and
- the extensive window-dressing of benefits that is found in the decisions and evaluations of most state and local infrastructure projects.

My inclination, after reviewing this evidence, is to refute the extremes of criticism as well as the promises of benefits from reforms. Undoubtedly, there have been major mistakes in transportation planning and undertakings. The over-building of railway lines in Western Canada early in this century, and of Mirabel airport in the second half of this century, are examples that come to mind. But every line of public and private venture contains successes and failures. The question is whether the mistakes in transportation are the exception rather than the rule. Canada and the United States could not have had as successful a half-century of economic

growth as they have had since the end of World War II if their transportation investments and operations were failures, or, at least, gross errors.

I am inclined to accept the arguments about the benefits of increased application of user-charges and earmarked taxes in transportation. New opportunities have arisen from new technologies. While not as pessimistic as Altshuler about their acceptability, I believe that there will be a good deal of inertia in extending user-charges in transportation. Furthermore, the efficiency improvements are not likely to be as large as Winston, Schultze and Small maintain.

I am more inclined to accept the Aaron, Schultze, Winston, and Musgrave judgement on the size of the economic returns to infrastructure investment than I am to accept the Aschauer judgement. But the line of analysis opened up by Aschauer, and explored by many others, has added to the evidence on the role of infrastructure on the economy as a whole, and on regions and sectors.

I acknowledge that a good deal of fakery, puffery, and wrongheadedness exists in the analysis and decision making on investment and management of infrastructure, including transportation. But a good deal of weight should be given to the other side of this argument. Much more careful and reliable benefit-cost analysis is done in these areas than was carried out 20 or 40 years ago. Computer-based analysis and information bases have enormously improved these aids to decision making. A highly trained and experienced profession of planners and evaluators has developed as a result.

PART B

TRANSPORTATION AND REGIONAL ECONOMIC DEVELOPMENT (WITH SPECIAL EMPHASIS ON PASSENGER TRANSPORTATION)

QUESTIONS AND CONCEPTS

In this part of the survey, the main issues are:

- Has transportation shaped the *regional pattern* of Canadian economic development?

- Has passenger transportation shaped the *regional pattern* of Canadian economic development? In what ways? Have the transportation activities been beneficial? Where they have not been fully effective, why not?
- Are transportation (particularly passenger transportation) projects and policies effective ways for improving the overall and regional features of Canadian economic development?
- What are the interactions between transportation and other factors affecting the regional features of Canadian economic development?
- While the interests in this survey are ultimately Canadian, much of the literature is based on U.S. experience. Because of similarities in economic history and geography, however, that literature provides lessons for Canada.

A pragmatic approach to the concept of “regions” seems best for this survey. The regions that are considered to be affected or affectable by transportation developments are: metropolises and their satellites; secondary cities; “heart-lands” and “hinterlands,” as geographers call them; provinces; and regions within provinces and states, singly or in groupings. The main concern of this survey is inter-regional, but the structure of modern economies requires consideration of intra- as well as inter-regional developments and the relationships between them.

TRENDS OF REGIONAL ECONOMIC DEVELOPMENT IN CANADA AND THE UNITED STATES

What have been the main regional patterns of Canadian and American economic development during the last half-century? Superficially, how have changes in transportation affected them? How have the patterns of economic development affected transportation?

CONSIDER CANADA FIRST

While the production and distribution of resource-based export staple commodities have become smaller shares of Canadian GDP than they were five and ten decades ago, they are still important activities. For the agricultural,

fishing, forestry and mineral products industries (oil, gas and derivatives), transportation continues to be a major factor. It is required mainly for the movement of commodities, although passengers who service these staple industries also require transportation. Location of extraction and processing activities continues to be mainly at the resource base, and these bases are spread across the country. Thus, Canada's resource industries and its transportation shape many of the country's regions. The decline of the share of the work force involved in agriculture has been particularly dramatic.

Secondary manufacturing increased and then fell back as a share of Canadian GDP, but continues to be a major group of activities. To a surprising degree, these industries are centred in the major metropolitan areas — particularly in and around Toronto and Montreal. Within these urban areas, manufacturing has moved from the city centres to the suburbs and satellite cities. (See the *Historical Atlas of Canada*, Volume III, Plates 7, 12, 13, 14, 51, 54, and 55.) While transportation costs do not appear to be as critical for the location or productivity of secondary manufacturing as they used to be, they are still important. Moreover, speed and reliability of transportation services have become more important for many secondary manufacturing industries.

Service Industries

The largest recent change in the structure of the Canadian economy, aside from the decrease in the agricultural work force, has been the growth of service industries. Aside from rural recreational activities, the bulk of service activities have located in and around metropolises. This is not surprising in view of the Grubel and Walker (1989) evidence that about half of service activities are business services of one kind or another. Transportation of people and interpersonal communications are major causes of the development of these parts of the service economy.

Health care and education have grown more rapidly than the GDP and the work force; these too are predominately located in and around metropolises. School consolidation and busing have increased the urban concentration of elementary and high school education.

The enormous increase in the participation of women in the labour force has affected the structure of the economy and has interacted with transportation. It has:

- intensified growth in ownership and use of private cars;
- intensified the substitution of household equipment for household labour; and
- been a catalyst for the rapid growth of fast-food outlets and commercial consumer services.

Regions

Cities outside the Canadian heartland have become ambitious to diversify their activities and outgrow regional servicing functions.

Despite a huge increase in productivity for the Canadian economy as a whole, poor and low-productivity regions persist. Newfoundland, much of the Maritime provinces, and parts of rural Quebec appear to not have improved their comparatively low productivity, slow growth, high unemployment and under-employment in relation to the Quebec-Ontario heartland, Alberta and British Columbia.

Regions in which Aboriginal peoples live continue to have notoriously poor economic performance, with a few exceptions. Even within prosperous provinces, pockets of poverty, some of which have a regional dimension, persist. Examples of these are areas in which declining manufacturing industries predominate, and areas in which the primary resource is declining.

The United States is witnessing many of these trends as well, and is seeing some significantly different trends, including:

- a movement of people to the sunbelt;
- a decline in numbers of unionized workers;
- industrial decline in the Great Lakes and midwestern regions; and
- a large and persistent illegal immigrant problem.

HAVE TRANSPORTATION ELEMENTS BEEN MAJOR FORCES SHAPING THE REGIONAL AND INTER-REGIONAL STRUCTURE OF THE CANADIAN AND UNITED STATES ECONOMIES?

Conflicting Opinions

Transportation gets mixed reviews in both countries with respect to regional economic development. Some writers attribute the economic development of a region to favourable transportation developments. Canadian examples of this include:

- the post-World War II highway links between central Canada and the U.S. interstate highway system that are largely responsible for the continental integration of the Canadian and U.S. manufacturing economies;
- the development of oil and gas pipeline networks and the post-World War II development of Alberta and Saskatchewan oil and gas resources; and
- the Quebec North Shore and Labrador Railway and the post-World War II development of the Labrador Trough.

Many writers attribute unsatisfactory regional economic development, in part, to inadequacies or mistakes in transportation developments and policies (for example the reduction in subsidization of Maritimes freight rates).

Some writers contend that transportation developments or policies could be the main force to improve the economic welfare of less-well-off regions. These regional transportation proposals are sometimes thought to be consistent with increased national GDP. Sometimes they are acknowledged to be mainly diversionary, but are advocated nevertheless. It is frequently contended that most regional transportation projects merely shift activity from one region to another, thereby *reducing* real national output. This view is given much weight by Adler, in Fromm (ed.) (1965); and by Hirschman (1958). Much of Hickling's *Primer* aims to prevent such outcomes.

It is important to recognize, nevertheless, that many transportation developments that improve national economic development necessarily have favourable effects on many regions. If investment in better pavements for highways greatly improves the efficiency of their use, it is likely that these same improvements will benefit most regions; trucking, after all, is

ubiquitous throughout Canada. If improved air traffic control technology would improve the national efficiency of air service, it is likely to do so in most regions, though not necessarily to the same degree. If real incomes produced or received increase in most regions, even though relative gaps are changed little, the demand for some kinds of transportation services will increase in most regions, though again, not necessarily to the same degree.

Thus, the issues of transportation and regional development have to be appropriately formulated. For example:

- Are national trends in transportation causing national economic development to be applied appropriately in various regions?
- Are some trends in transportation more effective for some regions than for others; and do the projects and policies reflect this?
- Do some problems in transportation, such as congestion, occur everywhere, although more severely in some regions than in others?
- Is transportation a more suitable agent of economic development in some regions than in others?

LITERATURE TO BE SURVEYED

The literature to be surveyed includes:

On regional development theory and policy: Adler, in Fromm (ed.) (1965); Eberts (1991); Fox (1990); Green (1971); Isard (1975); Kraft et al. (1971); McCann (1987); Norrie and Percy (1988); Savoie (1981 and 1986); Sitwell and Seifried (1984); Studnicki-Gizbert (1990); Sullivan et al. (1989); Vernon (1969); and Munnell (1990b).

On empirical regional and transportation literature: Deno (1988); Eberts (1991); Fox (1990); Friedlaender (1990); and Munnell (1990a).

THEMES IN THE LITERATURE

Among the literature on transportation and regional development, several themes appear repeatedly. The same concepts and measurements apply regionally as they do nationally for deciding on and evaluating

transportation and economic development. Transportation projects are rarely the main determinant of location or productivity of activities in a region. Except for resource industries, they are generally one element interacting with many other elements in determining regional economic development. If many other elements are not favourable to economic development in a region, a major transportation investment is not likely to be worth the cost.

The integration of transportation projects and policies and other regional economic development policies has seldom been done effectively. Indeed, in recent books on regional economics, it is surprising how little attention is given to transportation. In Savoie (1986), for example, there are three minor comments on transportation.

SOME DETAILS FOR VARIOUS KINDS OF REGIONS

Any consideration of transportation and economic development should deal with such diverse regions and inter-regional relationships as were noted above. They should deal both with regions with high employment, productivity and growth, and with those with low employment, productivity and growth. Consideration of regionalism and economic development should not be confined to the so-called "have-not" regions and the "poverty pockets."

During the 1950s and 1960s, the literature on economic development paid considerable attention to public investment and social overhead capital. Hirschman (1958) notes the political and social pressures in both developing and developed countries to scatter public investment over all regions in a country. Often the national economic development results were poor. Fox (1990) remarks on the limited empirical evidence that exists on the regional effects of infrastructure developments and policy, despite Hirschman's early interest in the subject.

Fox writes:

Application of research results to improve policy requires understanding how the productivity of infrastructure varies at different locations, and why. Presumably, a given marginal infrastructure investment shifts the production frontier out by different amounts across regions. Further, actual production may not respond to a shift in the frontier,

particularly if the economy was already operating inside the previous frontier. However, little empirical evidence exists on how productivity effects vary by region, despite longstanding conceptual arguments that considerable difference should be expected (for example, see Hirschman, 1958) and it is difficult to generalize the findings of what research exists. The lack of good guidelines on infrastructure's productivity in different geographic and economic environments is perhaps the most troublesome limitation imposed by the current base of empirical knowledge. (p. 32)

In reviewing transportation and regional development, this paper examines:

- large, mature, well-off regions;
- hinterlands, between-hinterlands and export markets;
- hinterland cities seeking diversification;
- passenger movements in diverse regions;
- have-not and have-less regions; and
- Canadian programs and regional inequalities.

In large, mature, well-off regions, transportation is one element of integrated development processes. Such regions generally have a well-developed transportation system with a high density of roads, railroads, and airport facilities. Kraft et al. (1971) make note of this when they remark that:

- On the whole, transport will not greatly influence the shape of future regional development in the United States, although under certain circumstances it may help a region to capture a fair share of decentralizing industry. In this respect, service considerations more than costs may determine the attractiveness of a particular transportation network.
- The diminishing role of transportation in influencing location choices should not be too surprising in a country where the network is very extensive already. (p. 35).

Similar conclusions apply to Canada now, although the Trans-Canada Highway network is less developed than the Interstate Highway System in

the United States. Transportation will have to adapt to changing demands, growth and technology.

As discussed in the general treatment of transportation and economic development, Western industrialized countries have, for decades, concentrated their economic activity in and around cities. These concentrations, however, differ in their degree of specialization. Porter (1990) emphasizes the clustering of related firms in certain cities.

Axiomatically, better-off regions have the wealth and income to maintain and improve transportation. However, judging by the pleas of U.S. governors for federal funds and for tax-exempt status for transportation bonds, the willingness of even the better-off states to be self-sustaining in infrastructure investment and maintenance is limited. (See U.S. Governors' Task Force Report, 1989; and Gramlich in Munnell (ed.), 1990.)

Economic development can be largely a self-sustaining process, although in the past it has depended on certain public goods, such as education, transportation and communications, and research and development.

Whether public policies on transportation and economic development should focus on increasing infrastructure, or improving its efficiency of design and use, or some mixture, is a moot point. The governors (1989) and many economists (Galbraith, 1991; Joint Economic Committee hearings, 1990) emphasize increased investment. As shown in Part A of this survey, the Congressional Budget Office, Schultze, Small and Winston emphasize improving the efficiency of investment and use. These arguments are being fought for in every better-off region in the United States. The situation in Canada is unclear, although highway authorities and municipal governments in well-off regions are pleading for more provincial and national financing of infrastructure investment.

It appears that the movement of people within and among the more highly developed regions in North America will continue to grow more rapidly than the real output of goods and services of the regions. This inference is drawn from Winston's 1985 survey. Even if much improvement in efficiency in investment and use takes place, eventually some increase in the stock of infrastructure capital (particularly in transportation) will be worthwhile. As Fosler commented in Munnell (ed.) (1990), economic and social pressures will support such increased investment sooner rather than later.

The major transportation problems in the well-off regions of Canada are likely to be congestion and maintenance of the infrastructure.

TRANSPORTATION AND ECONOMIC DEVELOPMENT IN THE HINTERLAND AND BETWEEN THE HINTERLAND AND EXPORT MARKETS

Transportation is central to staple product production and distribution of hinterland. If a new resource project is proposed in an area not served by transportation, then the development and cost of transportation is an integral part of the decision for the project. Examples of this include the Quebec, North Shore and Labrador Railroad; Pine Point Mine; North-East B.C. Coal; Alberta Pulp and Paper; and Beaufort Sea and Mackenzie Valley oil and gas. If the new resource project is already served by transportation, then the prospects for the project are improved.

Transportation has often led economic development and shaped its regional distribution (see Eberts, 1991; Fogel, 1964; Fox, 1990; Kraft et al., 1971; McCann, 1987; and Sitwell and Seifried, 1984). Canals, waterways, ports, railroads and roads are well-known examples of how this happens in the United States and Canada. More often than not, these transportation developments involved public investment.

The consensus in the literature is that transportation is a necessary but not sufficient condition for the economic development of areas that have not shared fully in the economic development of well-to-do countries. However, even if transportation is necessary, it is usually considered to be an unimportant element in regional development programs. Eberts (1991) writes:

... public infrastructure is more a necessary condition than a sufficient condition for economic development. While public infrastructure construction can provide local jobs, unless the project is of considerable size and ongoing, sufficient demand to sustain local economic development must come from other sources. Still, the question needs to [be] raised, "Would the investment have occurred without the transportation investment?" (p. 7) [He later asks] Does transportation induce economic development? The answer, gleaned from the current body of empirical research is a qualified yes. . . . Furthermore, the positive effects of transportation are seen at both the local and national level. [This remark appears to refer to the literature generated by Aschauer

et al.] If the effects were observed at only the local level, one would be inclined to ask whether transportation investment, and other infrastructure investment, is simply rearranging a fixed pool of resources, benefiting those regions that have attractive infrastructure investment at the direct expense of other regions. However, the observed national positive effects allays these concerns to some extent. The ability to stimulate economic growth depends on the economic state or development of a region. (pp. 21–22); see also Sullivan (1989.)

Hinterland Cities

Hinterland cities seek diversification of their economies; they want to be something more than service centres for their hinterlands. Poor or expensive transportation services are alleged by some writers to be a major limitation on such diversification (Studnicki-Gizbert, 1990; and McCann, 1987).

Governments have tried diversifying in ways that avoid transportation becoming a decisive impediment, such as concentrating on high-tech, high-value goods and services, and on intellectual capital. These have been the purpose of the Alberta diversification programs and the federal government's Western Diversification Program.

Residents and governments of the Maritimes and the Prairies have made repeated complaints about transportation impediments to their diversification and their opportunities to penetrate markets in central Canada and the United States.

Moving People

While much of the attention on transportation and regional economic development has been on the movement of goods, satisfactory services for the movement of people are also essential.

Movements of people arise from migrations, seasonal labour flows, trade, information exchange, management practices, and educational activities.

Land that is unsuitable for agriculture is sometimes the most attractive for cottages, camping and tourism. Familiar examples of this include the Canadian Shield north of Toronto; the Laurentians; the Rockies; and the Manitoba-Ontario-United States border country. Effective transportation services (such as Highway 400 north of Toronto) are essential in these areas.

One of the clearest examples of the impact that transportation has on economic development and movement of people is the highway development in Newfoundland. Newfoundlanders used to rely on boat connections among the outports and between the main centres and the outports; today they rely on cars and trucks. The Trans-Canada Highway in Newfoundland has brought about profound changes in the province's social and economic life.

The "Have-nots" and the "Have-less" Regions

One of the main concerns of regional development in Canada has been improving the lot of the "have not" and the "have-less" regions.

Even if policies and programs contribute little to or even decrease national product, most highly developed Western nations do not take a *laissez-faire* approach to the people and institutions in their "have-less" regions. Regional aspects of economic development have been of concern in less-developed countries too, and involve economic, political, social and equity considerations (see Bothwell et al., 1989; Courchene, 1981; Economic Council of Canada, 1977 and 1980; Friedman and Alonso, 1964; Green, 1971; Hickling, 1990; Hirschman, 1958; Kraft et al., 1971; Lithwick, 1978; Savoie, 1981 and 1986; Scott, 1978; and Simeon and Robinson, 1990).

It is usually best to first consider equalization transfers to people and governments, and then to consider regional development *per se*, such as projects and policies. This survey is concerned with regional development projects and policies.

Regional Transportation Policies

Transportation often has important regional development effects, even when it is not integrated into policies and programs that are more commonly called "regional." In Canada, transportation programs and policies of major regional importance have included:

- building railways and roads to serve less-well-off regions, with considerable public investment and subsidy;
- subsidies for transportation services, such as Crow rate, feed grain movements, and Maritime freight rates; and

- development of proposals for region-serving transportation, such as the gas pipeline to Vancouver Island, and the proposed Maritime extension of the gas pipeline from Eastern Canada.

Studnicki-Gizbert (1990) examines the different approaches to subsidies in Canada; he is highly critical of several of the methods used, and points to inappropriate incentives and inefficiencies.

Regional considerations also arise in transportation regulation and deregulation, such as:

- trucking deregulation, in assurance of reliable, economical common-carrier service for remote areas;
- airline deregulation, in proposals for assurance of reliable and economical service for remote areas; and
- branch line closings, in consideration of the adequacy of remaining transportation services.

Substantial national and provincial regional development programs have been implemented since the latter 1950s, many of which are now summarized in the Fall 1990 Canadian Labour Market and Productivity Centre (CLMPC) *Quarterly Labour Market and Productivity Review*. The best reviews are in the books by Donald Savoie, and are listed in the CLMPC article.

These programs, while they had an infrastructure component, did not contain much transportation activity. It may be worthwhile for the Royal Commission to examine some of the Federal programs under Department of Regional Economic Expansion (DREE), Department of Regional and Industrial Expansion (DRIE), and successors, and federal-provincial agreements under these programs, to determine the transportation components, both for goods and for people.

The programs were laden with subsidies to private investment, either to draw in activities from outside the region or to support local initiatives. In recent years, the latter strategy has come into more favour. These Canadian programs have been plagued by local "pork barrelism," and nearly every politician usually promotes, or at least acquiesces to, such activities. Every

town in the Maritimes seeks its share of development funds. Sub-regions within well-to-do provinces seek their share too. Growth-pole ideas were badly abused in Canada.

Reducing Regional Disparities

The conventional wisdom among analysts of Canada's regional policies is that disparities in productivity and produced income have *not* been reduced by such policies. It is agreed, however, that disparities in *received* income have been substantially reduced, mainly due to federal-provincial transfers to provinces and to people, including equalization, Established Programs Financing (EPF), sharing of Canada Assistance Plan (CAP), and the transfers to people through Old Age Security/Guaranteed Income Supplement (OAS/GIS), Unemployment Insurance Commission (UIC), and Canada Pension Plan/Quebec Pension Plan (CPP/QPP). Not all of these programs are regional *per se*; indeed few of them have specific regional considerations built in. However, they have been the main forces in reducing the regional disparities in Canada, with respect to income received by persons and household.

Movement of People

The improved relative *income-received* position of persons in the have-less regions has been accompanied by increased movement of people in all categories, and therefore has called on improved transportation services. An excellent illustration of this is the greatly increased movement of people within the 100-mile radius around Halifax as compared with one or two decades ago.

CONCLUSIONS REGARDING TRANSPORTATION AND ECONOMIC DEVELOPMENT

Many transportation and passenger transportation elements that have contributed to national and international economic development have been shared among the regions, though not always equally:

- good, paved highway systems have been developed in centres where there is significant population density, and for connections between these centres;

- private ownership of cars and light trucks has become considerable in all regions;
- reliable rapid air passenger service, and air traffic control is available throughout the country, including remote areas; and
- oil and gas pipelines serve most regions that have significant population density.

On the negative side, all regions have experienced a deterioration in railway passenger services and little improvement in intercity bus services. The one exception is the improvement in commuter train services for Toronto and environs.

Transportation's Role in Regional Development

Participation in these transportation-economic development changes has led to investment in transportation in many regions. The literature contends that overbuilding of highways in rural areas occurs more often in the U.S. than it does in Canada.

Some region-specific economic developments are dependent on transportation facilities. Other economic developments have been important contributors in the development of multi-purpose transportation facilities. For example, many resource-sector economic developments have been dependent on road, railroad, and air services. As another example, tourism has often been a deciding factor in the decision to turn a local or regional road into a major highway.

In mature, better-off regions, the main transportation problems are congestion and maintenance. Such regions are capable of self-financed transportation developments, with some allowance for spillover.

When it comes to boosting the economic development of less-well-off regions, transportation and general economic development policies have not been well integrated. Transportation projects appear to be decided by authorities that are distinct from and often out of touch with development authorities and *vice versa*. Admittedly, there are some infrastructure components in general programs concerning regional disparities, but the lack of integration is the rule rather than the exception.

Transportation is often thought to be a minor element in general regional development policies. Paradoxically, advocates of transportation developments often, at the same time, allege regional development benefits in support of their projects.

A couple of decades ago, airports were widely touted as being growth poles (for example, Mirabel and Pickering). These expectations have been severely discounted since then.

On balance, the consensus among transportation economists is that there is a bias toward exaggeration in alleging that transportation projects produce regional economic development benefits.

PART C

TRANSPORTATION AND TOURISM

INTRODUCTION

Intercity movements of people can usually be divided into those for business and those for non-business purposes. (There is some overlap, such as travel to conventions.) Non-business purposes are the aggregation of travel for visiting friends and relatives, for pleasure and for personal reasons.

The principal concern in this part of the survey is travel for non-business purposes, which will be called "tourism." The Royal Commission's Interim Report shows that, in 1988, tourism comprised 83 percent of domestic intercity trips. The report also shows that:

- Canadians travel primarily within their own provinces;
- most trips are for short and medium distances;
- international travel is increasing;
- the transportation mode that travellers choose depends on the length of trip. For short trips, cars are used 95% of the time, air almost never, bus about 3% of the time, and rail about 1%. Use of cars falls to 81% for trips between 320 and 800 kilometres in one-way distance, while air jumps into

second place with 12% for such trips. Bus and rail increase their shares to 4% and 3%, respectively. As trip lengths increase further, air's share increases and the shares of the other three modes decrease;

- Canadians travel primarily by car;
- use of a personal vehicle consumes approximately 91 cents of every consumer transportation dollar; and
- air travel is the most popular mode of public transportation. For long trips, Canadians tend to choose air travel. In 1988, 29% of business trips and 3% of pleasure trips were taken by air.

This section of the survey is concerned with the causal links between transportation and an aspect of economic development, tourism.

Here are the critical questions:

- Is there convincing evidence that transportation projects, policies and operations have shaped tourist activities? How? Why? Have the results been relatively efficient?
- Does that evidence indicate important forces of future development of tourism? Which of these forces arise from transportation? Does the evidence point to policy options for current or future decisions?
- Is there convincing evidence of causal links from tourism to transportation projects, problems, policies and decisions?

Transportation developments clearly cause tourism developments, and *vice versa*. The development of large, fast, economical jet passenger aircraft has led to long-distance tourism. Tourism in the Canadian Shield has led to demands for and provision of better access roads. *Precise analysis of these relationships, and in particular of their changes over time, is difficult.*

Particularly important issues for transportation policy are the response of travel (both personal and business) to changes in the elapsed time and cost of trips, and the incomes and characteristics of travellers. (Economists call these measurements "elasticities of demand.")

The literature examined in this survey includes: Bothwell et al. (1989); Gillen and Oum (1981); Grubel and Walker (1989); *Mathematica* (1966 and 1967); McCann (1987); Oum and Gillen (1983); Oum, Waters and Yong (1990); Palmer (1988); Scarfe and Krantz (1988); Watson (1988); and Winston (1985).

TRANSPORTATION SUPPLY

A number of transportation developments have increased tourism in North America in general, and in Canada in particular. These have included:

- the enormous increase in private car ownership and operation. (Bothwell et al., 1989 show that mass ownership of private cars in Canada occurred *after* World War II);
- the development of a highway system of paved roads. (Even more than the mass ownership of private cars in Canada, this is a post-World War II phenomenon in Canada [Bothwell]. In the United States the most visible indicator of this is the near-completion of the Interstate Highway System by the early 1970s. Canada has its Highway 401 and Trans-Canada Highway from St. John's to Victoria);
- the development across Canada of a network of civilian airports, with navigation and air traffic control services for dense movement of aircraft, by day and night, in all weather. (The nucleus of the system was given impetus from the wartime Commonwealth Air Training Program, as mentioned in the Bothwell and Kilbourn biography of Howe);
- the availability of reliable, fast, large, economical passenger aircraft;
- the availability of cheap motor fuels; and
- the proximity to Canada's main metropolises of the recreational potential of the Canadian Shield. Land that was virtually useless for agriculture is now the home of vast supplies of rivers and lakes, cottages, campsites, ski hills and so on.

IMPORTANCE OF TOURISM

Though the available data is incomplete, it contains many indicators of the important place that tourism plays in the lives of Canadians. Watson (1988), for example, shows the growing household ownership of recreation

equipment, such as vacation homes, camping equipment, boats, snowmobiles, skis, cycles, motorbikes, and adult-sized bicycles (Table 33). Watson also:

- provides summary data on family expenditures on various forms of recreation or recreational equipment . . . [and notes that] consumers' spending on entertainment and recreation increased from 4.0 percent of GNP and 6.3 percent of total personal expenditure in 1961 to 6.0 percent of GNP and 10.1 percent of personal expenditure in 1983. It is not at all surprising, therefore, that most of the items tracked in Table 34 have also increased more rapidly than total personal expenditure, which rose by just over three and a half times (in current dollars) between 1969 and 1984. For instance, spending on vacation homes rose more than fourfold; spending on holiday lodging more than four and a half times; purchases of recreational vehicles more than five times; purchases of bicycles almost five times. . . . (pp. 78–80)

Watson also comments on the growth in the use of national parks, campground activities, national historic parks and provincial parks — another set of indicators of growth in tourism more than in proportion to population and income (Table 36).

The tourism into and out of Canada is large, though still much smaller than the domestic aspect.

MODES OF TOURING

Modes of transportation for tourism have already been summarized from the Royal Commission's Interim Report: the mode for short and medium trips is personal car, with air travel being of increasing importance for medium- to long-distance trips.

Palmer (1988) provides useful data on passenger trips by bus and rail. Between 1975 and 1985, the trend has been for an absolute decline in intercity common carrier bus trips, but not much decline in comparable rail trips. Palmer argues that, "Regardless of some qualms about these data, it appears that passenger rail service may be a normal good, whereas passenger bus service is more than likely an inferior good [p. 46]."⁸

ELASTICITIES OF DEMAND

Trends in trips, expenditures and income suggest that the income elasticities of demand for tourism activities as a group are more than one. For example, a 10 percent increase in income will lead to more than a 10 percent increase in expenditure on tourism. (The income elasticity of demand of individuals, or, for that matter, for groups of individuals for particular elements of tourist activity, may be well below unity without contradicting the conclusion about aggregates.)

However, analyzing demand and family expenditure is difficult, and full of complexities. The classic literature on elasticity analysis is not reviewed here. Note is taken, however, of the review article by Winston (1985). For demand, Winston reports favourably on models that examine the responses both to elapsed time of trip and cost (price) for various modes and trips. Estimates of price and service-time elasticities are reported.

Winston notes:

These coefficients can be used to calculate estimates of price and service time elasticities of demand and decisionmakers' value of travel time. . . . In contrast to elasticity estimates for urban passenger transportation, service-time elasticity estimates for intercity bus and rail transportation tend to be larger than the price elasticity estimates. Furthermore, their large magnitude . . . indicates that reductions in service times could be significantly effective in increasing rail and bus-market share. Generally, the cost and service-time elasticities for air and auto are inelastic. This is not too surprising in view of the fact that these modes already possess a relatively large share of the United States' intercity travel market. (pp. 73–75)

TRANSPORTATION RESPONSES TO TOURIST DEMAND AND SUPPLY

Some transportation developments take place in response to tourist use and demands. They may follow some developments and then induce further tourist development. Particularly good examples of this in Canada are Highway 400, north of Toronto; the Laurentian Autoroute, north of

Montreal; the Calgary–Banff–Lake Louise–Jasper–Edmonton highway and rail service; the early development of a good paved highway from Winnipeg to Kenora; and the Cabot Trail on Cape Breton Island.

Ancillary developments are also important to the transportation-tourism mix, and include: airport hotels; highway service centres; multi-purpose resort complexes; ski hills and lifts; marinas; and parks and camp grounds.

A particularly important activity for long-distance and foreign travel has been the development of reservation systems, tour packagers and wholesalers, and an industry of travel agencies. These serve both tourist and business travel.

ENDNOTES

1. Kindleberger (1965), pp. 189–90. [Emphasis mine.] In more recent literature, the slowdown in economic development is sometimes attributed to the reduction in the redundancy of labour, with the consequent increase in real wages and reduction in real profits. See Cornwall (1977) and Boltho (1988).
2. On infrastructure and location of industry, however, see Munnell (1990) in the volume she edited for the Federal Reserve Bank of Boston.
3. Some of these activities may still be *natural monopolies* in the technical sense. Their activities may still be operating in regions where declining average costs are large relative to the total (narrowly defined) market. Even in these cases, however, there is generally more competition for the activity from close substitutes.
4. Kindleberger (1965) pp. 157–58.
5. These observations have often focussed on transportation, with the same alleged cause-effect relationships.
6. Fox (1990): “The Contribution of Infrastructure Investments to Growth: A Review of the Literature.”
7. Essentially the same ideas are found in Mohring and Harwitz (1962), but are developed by them in a more technical way.
8. For a normal good, the income elasticity of demand is positive, though it could be large or small. For an inferior good, the income elasticity of demand is negative; that is, an increase in income, other things being equal, induces a decrease in the quantity demanded.

REFERENCES

- Aaron, Henry J. 1990. "Discussion." In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell.
- Abramovitz, M. 1952. "Economics of Growth." In *A Survey of Contemporary Economics*. Edited by B. F. Haley. Vol. II. Chicago: Irwin.
- Adler, Hans A. 1965. "Economic Evaluation of Transport Projects." In *Transport Investment and Economic Development*. Edited by Fromm.
- Altshuler, Alan A. 1990. Discussion of Winston, "How Efficient is Current Infrastructure Spending and Pricing?" In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell.
- American Association of State Highway Officials. 1977. *A Manual on User-Benefit Analysis of Highway and Bus-Transit Improvements*. (Cited in Hickling, December 1990.)
- Apogee Research, Inc. October 1989. *Current Literature on Highway Investment and Economic Development*. U.S. Federal Highway Administration. (Cited in Hickling, December 1990.)
- Arrow, Kenneth J. 1969. "The Social Discount Rate." In *Cost-Benefit Analysis of Manpower Policies*. Edited by C. G. Somers and W. D. Wood. Kingston, Ontario: Queen's University, Industrial Relations Centre.
- Aschauer, David A. 1988. "Government Spending and the Falling Rate of Profit." *Economic Perspectives: A Review of the Federal Reserve Bank of Chicago* 12.
- . 1989a. "Is Government Spending Productive?" *Journal of Monetary Economics* 23.
- . 1989b. "Does Public Capital Crowd Out Private Capital?" *Journal of Monetary Economics* 24.

- . 1989c. "Public Investment and Productivity Growth in the Group of Seven." *Economic Perspectives: A Review of the Federal Reserve Bank of Chicago*, 13, 5.
 - . 1990a. "Why is Infrastructure Important?" In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell. (Includes important comments by J. Aaron Henry and Richard A. Musgrave.)
 - . 1990b. "Highway Capacity and Economic Growth." *Economic Perspectives: A Review of the Federal Reserve Bank of Chicago*, September-October 1990.
- Baldwin, John. 1989. *Regulatory Failure and Renewal*. Economic Council of Canada. Ottawa: Supply and Services Canada.
- Barro, Robert J. 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy* 81.
- . 1990. "Government Spending in a Simple Model of Endogenous Growth." *Journal of Political Economy* 98 (October 1990).
 - . 1989b. "A Cross-country Study of Growth, Saving, and Government." Unpublished. (Cited in Aschauer, 1990.)
- Batchelder, James H. 1979. *Application of the Highway Investment Analysis Package*. U.S. Transportation Research Board Record. (Cited in Hickling, December 1990.)
- Bird, Richard. 1976. *Charging for Public Services*. Canadian Tax Foundation. Toronto.
- Boadway, Robin W. 1985. "Public Finance," in *The Canadian Encyclopedia*, second edition, Vol. III. Edmonton: Hurtig Publishers, pp. 1781–83.
- Boltho, Andrea, ed. 1988. *The European Economy, Growth and Crisis*. Oxford: Oxford University Press.

- Bos, Deiter. 1985. "Public Sector Pricing," *Handbook of Public Economics*, Vol. 1, Edited by A. J. Auerbach and M. Feldstein. Amsterdam: North-Holland.
- Bothwell, Robert et al. 1989. *Canada Since 1945: Power Politics and Provincialism*. Revised Edition. Toronto: University of Toronto Press.
- Brander, J., B. Cook and J. Rowcroft. 1988. "Auctioneering Runway Landing Slots: Long-Run Implications." *Transportation Research Forum*.
- Buchanan, James M. and Gordon Tullock. 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.
- Canadian Labour Market and Productivity Centre. 1990. "A Review of Regional Economic Development Policy in Canada." *Quarterly Labour Market and Productivity Review*. Fall 1990.
- Careless, J. M. S. 1979. "Metropolis and Region." *Urban History Review* 18. (Cited in *The Canadian Encyclopedia*, 1985.)
- Caves, R. E., M. E. Porter and M. Spence, with J. T. Scott. 1980. *Competition in the Open Economy: A Model Applied to Canada*. Cambridge, Mass.: Harvard University Press.
- Choate, Pat. 1984. "Whatever Happened to Infrastructure?" Paper prepared for presentation at "For a Growing America" — a *Centre for Policy Research Conference*.
- Coase, Ronald. 1960. "The Problem of Social Cost." *Journal of Law and Economics*, 2 (October 1960).
- . 1974. "The Lighthouse in Economics." (Cited in *The Economist*, February 23, 1991, p. 73.)
- Cornwall, John. 1977. *Modern Capitalism, its Growth and Transformation*. London, U.K.: Martin Robertson.

Costa, Jose da Silva, Richard W. Ellson and Randolph C. Martin. 1981. "Public Capital, Regional Output and Development: Some Empirical Evidence." *Journal of Regional Science* 27.

Courchene, Thomas J. 1981. "Avenues of Adjustment: The Transfer System and Regional Disparities." *Canadian Public Policy* 7, 4.

Cubukgil, Adil and Richard Soberman. 1986. *The Need for Rail Passenger Services and Opportunities for Modal Substitution*. Study funded by the Research Branch of the Canadian Transport Commission, University of Toronto-York University Joint Program in Transportation.

Dean, Durand, Fallon and Holder. 1989. *Saving Trends and Behaviour in OECD Countries*. Paris: OECD.

Deno, Kevin T. 1988. "The Effect of Public Capital on U.S. Manufacturing Activity: 1970 to 1978." *Southern Economic Journal* 55.

Dobell, A. R. and H. E. English. 1988. *Canada's Trade in Services: An Overview*. Institute for Research on Public Policy report to Canada, Department of Regional Expansion under its Service Industries Sector Study Program. Contains extensive bibliography on this subject.

Dornbusch, Rudiger, Paul Krugman and Yung Chul Park. 1989. "Meeting World Challenges: U.S. Manufacturing in the 1990s." Rochester, N.Y.: Eastman Kodak Company, Communications and Public Affairs.

Eberts, Randall W. 1991. *Transportation and Local Economic Development*. Federal Reserve Bank of Cleveland. Paper prepared for Transportation Research Conference, Washington D.C., January 16, 1991.

Eckstein, Otto. 1958. *Water Resource Development*. Cambridge, Mass.: Harvard University Press.

Economic Council of Canada. 1977. *Living Together: A Study of Regional Disparities*. Ottawa: Supply and Services Canada.

—. 1980. *Newfoundland: From Dependency to Self-Reliance*. Ottawa: Supply and Services Canada.

- . 1981. *Reforming Regulation*. Ottawa: Supply and Services Canada.
- . 1986. *Minding the Public's Business*. Ottawa: Supply and Services Canada.
- Economist*. September 22, 1990. "Nothing to Lose but Its Chains. A Survey of World Trade." (Excellent survey of theory, practice and policy.)
- . March 9, 1991. "Schools Brief. The Next 100 Years."
- Eno Foundation for Transportation, Inc. 1990. *Transportation in America*. Washington, D.C. (Annual issues with updates.)
- Federation of Canadian Municipalities. (On infrastructure requirements; mentioned in Swimmer paper.)
- Feldstein, Martin S. 1975. "Wealth Neutrality and Local Choice in Public Education." *American Economic Review*, March 1975.
- Fogel, Robert W. 1964. *Railways and American Economic Growth: Essays in Economic History*. Baltimore: Johns Hopkins Press.
- Forkenbrock, David J. 1990. "Do Highway Investments Spur Economic Development?" First Annual Transportation Research Conference, Center for Transportation Studies, University of Minnesota. (Cited in Hickling, 1990.)
- Fox, William F. 1990. "*The Contribution of Infrastructure Investments to Growth: A Review of the Literature*." Draft paper. The University of Tennessee. (Extensive bibliography, particularly on recent empirical work.)
- Freidman, John and William Alonso. 1964. *Regional Development and Planning: A Reader*. Cambridge, Mass.: M.I.T. Press. (Cited in Kraft et al. 1971 as a useful survey.)
- Friedlaender, Ann F. 1990. "Comment: How Does Public Infrastructure Affect Regional Economic Performance." In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell.

Fromm, Gary, ed. 1965. *Transport Investment and Economic Development*. Washington, D.C.: The Brookings Institution Transport Research Program.

Fuller, W. 1990. Cited in U.S. Department of Transport, *Moving America: A Look Ahead to the 21st Century*.

Galbraith, John Kenneth. January 1991. "Owning Up to the Recession." *Harpers March* 1991 (being a statement before the Senate Committee on Labor and Human Relations.)

Gaudry, Marc J. I. 1990. "Three Families of Mode Choice Models Applied to Intercity Travel Demand with Aggregate Canadian Data." Queen's University: John Deutsch Institute for the Study in Economic Policy. Discussion Paper No. 13.

Gay, Robert S. 1991. "The Investment Boom: End of a Cycle or a New Era?" In *Global Perspectives*. Edited by Morgan Stanley.

Gillen, D. W. and T. H. Oum. 1981. "The Structure of Intercity Travel Demands in Canada: Theory, Tests and Empirical Results." *Transportation Research*.

Gillen, D. W., T. H. Oum and M. W. Tretheway. 1988. "Entry Barriers and Anti-Competitive Behaviour in a Deregulated Canadian Airline Market." *International Journal of Transport Economics*.

—. 1988. "Airport Pricing Principles: An Application to Canadian Airports." *Transportation Research Forum*. (A clear and elegant paper on alternative pricing principles, emphasizing social marginal cost pricing, Ramsey quasi-optimal pricing, and peak/off peak pricing.)

Glaister, Stephen. 1981. *Fundamentals of Transport Economics*. Oxford: Basil Blackwell. (Cited in Hickling.)

Gomez-Ibanez, Jose A., John R. Meyer and David E. Luberoff. 1990. "What are the Prospects for Privatizing Infrastructure? Lessons from U.S. Roads and Solid Waste." In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell. (Also included are important comments by Sir Alan A. Walters and Gail D. Fosler.)

- Gramlich, Edward M. 1990. "How Should Public Infrastructure Be Financed?" In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell. (Also included are important comments by Rudolph G. Penner and James M. Poterba.)
- . 1990. *A Guide to Benefit-Cost Analysis*. Englewood Cliffs, N.J.: Prentice Hall.
- Green, Alan G. 1971. *Regional Aspects of Canada's Economic Growth*. Toronto: University of Toronto Press.
- Grubel, Herbert G. and Michael A. Walker. 1989. *Service Industry Growth*. In *The Economics of the Service Sector in Canada*. Vancouver: The Fraser Institute.
- Hahn, F. H. and R. C. O. Matthews. 1964. "The Theory of Economic Growth: A Survey. *Economic Journal*, December 1964.
- Harberger, Arnold C. 1969. "Discussion: Professor Arrow on The Social Discount Rate." In *Cost-Benefit Analysis of Manpower Policies*. Edited by G. C. Somers and W. D. Wood. Kingston: Queen's University, Industrial Relations Centre.
- Harris, Richard G. 1985. *Trade, Industrial Policy and International Competition*. Volume 13 of the research studies of the Royal Commission on the Economic Union and Development Prospects for Canada (the Macdonald Commission). Toronto: University of Toronto Press.
- Helliwell, J. F., Mary MacGregor and Tim Padmore. 1986. "Economic Growth and Productivity in Canada, 1955-90." In *Economic Growth: Prospects and Determinants*. Edited by John Sargent. Volume 22 of the research studies of the Royal Commission on the Economic Union and Development Prospects for Canada (the Macdonald Commission). Toronto: University of Toronto Press.
- Hickling Corporation. October 1990. *The Economic Impact of Alternative Runway Improvement Possibilities for Lester B. Pearson International Airport, Toronto*. Transport Canada. (Cited in Hickling, December 1990.)

—. December 1990. *A Primer on Transportation and Economic Development*. Second draft of NCHRP Report 342. (Prepared in association with Charles River Associates and Christensen and Associates.) Washington, D.C.: Transportation Research Board, National Research Council.

Higgins, Benjamin. 1983. "From Growth Poles to Systems of Interaction in Space," *Growth and Space* 14, 4 (Fall 1983).

Hirschman, A. D. 1958. *The Strategy of Economic Development*. New Haven: Yale University Press.

Hood, Wm. C. and A. D. Scott. 1957. *Output, Labour and Capital in the Canadian Economy*. Ottawa: Queen's Printer. (Study for the Royal Commission on Canada's Economic Prospects.)

Hotelling, Harold. 1959. "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates." *Readings in the Economics of Taxation*. Homewood, Illinois: Irwin.

Hunter, Holland. 1961. "Resources, Transportation and Economic Development." In *National Resources and Economic Growth*. Edited by Joseph J. Spengler. Washington, D.C.: Resources for the Future.

Hurley, William J. and Joseph Jones. 1990. *The Feasibility of High-Speed Rail in Canada*. Discussion Paper No. 14. Kingston: Queen's University, John Deutsch Institute for the Study of Economic Policy.

Isard, Walter. 1975. *Introduction to Regional Science*. Englewood Cliffs, N.J.: Prentice Hall.

Joint Economic Committee, United States Congress. July 1989. "Public Investment in Infrastructure." (Reference in Winston 1991 re testimony of economists on infrastructure.) (Important statements and written submissions to committee by Alan S. Blinder, Jack A. Meyer, Donald H. Staszheim and James Tobin, all supporting the view that substantial under-investment in public investment in human and physical infrastructure exists in the United States.)

- Jones, Ronald W. and Henryk Kierzkowski. 1988. *The Role of Services in Production and International Trade: A Theoretical Framework*. Halifax: The Institute for Research on Public Policy.
- Jorgenson, Dale W., Frank M. Gollop and Barbara M. Fraumeni. 1987. *Productivity and U.S. Economic Growth*. Cambridge, Mass.: Harvard University Press.
- Kaldor, N. 1967. *Strategic Factors in Economic Development*. Ithica, N.Y.: Cornell University Press.
- Keeler, Theodore E. and J. S. Ying. 1988. "Measuring the Benefits of a Large Public Investment: The Case of the U.S. Federal-aid Highway System." *Journal of Public Economics* 36.
- Kennedy, Paul. 1987. *The Rise and Fall of the Great Powers*. New York: Random House.
- Kerr, Donald, Deryck W. Holdsworth and Geoffrey Matthews. 1990. *Historical Atlas of Canada. Vol. III: Addressing the Twentieth Century*. Toronto: University of Toronto Press.
- Kindleberger, Charles P. 1965. *Economic Development*. Second edition. New York: McGraw-Hill.
- . 1988. *Europe's Post-war Growth: The Role of Labor Supply*. Cambridge, Mass.: Harvard University Press.
- Knight, F. H. 1924. "Some Fallacies in the Interpretation of Social Cost." *Quarterly Journal of Economics* 38.
- Kraft, Gerald, John R. Meyer and Jean-Paul Valette. 1971. *The Role of Transportation in Regional Economic Development*. (A Charles River Associates research study.) Lexington, Mass.: Lexington Books. (This book has an excellent bibliography of the theoretical and empirical literature up to the end of the 1960s, including the classic articles by Brian Berry, Edward L. Ullman, Bertil Ohlin, Albert O. Hirschman, Charles M. Tiebout, Holland Hunter, A. J. Bone and Martin Wohl, and Paul A. Samuelson.)

- Krutilla, John V. and Otto Eckstein. 1958. *Multiple Purpose River Development*. Baltimore: Johns Hopkins Press.
- Kuznets, S. 1963. "Comments." In *The Economics of Takeoff into Sustained Growth*. Edited by W. W. Rostow. London: Macmillan.
- Landau, Ralph. 1988. "U.S. Economic Growth," in *Scientific American* 256, 6 (June 1988).
- . 1990. "Capital Investment: Key to Competitiveness and Growth," in *The Brookings Review*, Summer 1990.
- Lewis, David, Daniel Hara and Joseph Revis. 1988. *The Role of Public Infrastructure in the 21st Century*. Special Report 220, Transportation Research Board. (Cited in Hickling, 1990.)
- Lewis, W. A. 1955. *The Theory of Economic Growth*. London: Allen and Unwin.
- Lithwick, N. H., ed. 1978. *Regional Economic Policy: The Canadian Experience*. Toronto: McGraw-Hill Ryerson.
- Mackintosh, W. A. 1964. *The Economic Background of Dominion-Provincial Relations*. Toronto: McClelland and Stewart. (Reprint of study for the Rowell-Sirois Commission.)
- Maddison, A. 1987. "Growth and Slowdown in Advanced Capitalist Countries: Techniques of Quantitative Assessment," in *Journal of Economic Literature* (June 1987).
- . 1988. "Ultimate and Proximate Growth Causality: A Critique of Mancur Olson on *The Rise and Decline of Nations*." In *Scandinavian Economic History Review* No. 2, (1988). (Cited in Maddison, *The World Economy*.)
- . 1989. *The World Economy in the 20th Century*. Paris: OECD Development Centre Studies.
- Markusen, James R. 1986. *Trade in Producer Services: Issues Involving Agglomeration Economies, Human Capital and Public Inputs*. A discussion paper. Halifax: Institute for Research on Public Policy.

- Mathematica*. Ed. 1966, 1967. *Studies in Travel Demand*. 3 volumes.
Princeton: Mathematica. (Cited in Kraft 1971 as an important study.)
- McCann, L. D. 1987. *Heartland and Hinterland, A Geography of Canada*.
2nd ed. Toronto: Prentice Hall.
- McFarland, William F. January 1987. *Ranking Highway Construction
Projects: Comparison of Benefit-Cost Analysis with Other Techniques*.
Texas Transportation Institute. (Cited in Hickling.)
- McFarland et al. November 1989 draft. *Economic Analysis of Transportation
Expenditures: A Literature Review*. Federal Highway Administration.
147 References. (Cited in Hickling, 1990.)
- McRae, J. R. 1982. *Canada-U.S. Trade in Transportation Services*. Institute
for Research on Public Policy.
- . January 1989. *Canada-United States Trade in Transportation Services*.
Halifax: Institute for Research on Public Policy.
- Meade, J. E. 1963. *A Neo-classical Theory of Economic Growth*. 2nd ed.
Fair Lawn, N.J.: Oxford.
- Melvin, James R. 1988. *The Role of Services in a Small, Open, Regional
Economy*. Halifax: Institute for Research on Public Policy, October 1988.
- Mohring, Herbert and Mitchell Harwitz. 1962. *Highway Benefits: An
Analytical Framework*. Evanston, Ill.: Northwestern University Press.
(Referred to in Winston 1991 as a classic on efficient transportation
infrastructure policy.)
- Montgomery, W. David. 1989. *Lessons From the Past, Opportunities for the
Future: The Changing Role of Public Transportation in Economic Growth*.
Colloquium on the Nation's Infrastructure Policy, November 17, 1989.
Washington, D.C. (Cited in Hickling.)
- Morrison, Steven. 1983. "Estimation of Long Run Prices and Investment
Levels for Airports Runways," in *Research in Transportation Economics*.
Greenwich, Conn.: JAI Press.

Morrison, S. and C. Winston. 1989. "An Econometric Analysis of the Demand for Intercity Passenger Transportation." *Research in Transportation Economics*. Greenwich, Conn.: JAI Press.

—. May 1990. "The Dynamics of Airline Pricing and Competition." *The American Economic Review, Papers and Proceedings*.

Munnell, Alicia H., ed. 1990a. *Is There a Shortfall in Public Capital Investment?* Proceedings of a Conference sponsored by the Federal Reserve Bank of Boston. (Main papers listed separately in this bibliography.)

—. 1990b. "How Does Public Infrastructure Affect Regional Economic Performance?" (Munnell's own paper in Boston Federal Reserve volume. Also includes important comments on this paper by Charles R. Hulten and Anne F. Friedlaender.)

—. 1990c. "Why Has Productivity Growth Declined? Productivity and public investment." *New England Economic Review*, January-February. (Cited in Aschauer 1990.)

Murphy, Kevin M., Andrei Shleifer and Robert W. Vishny. 1989. "Industrialization and the big push." *Journal of Political Economy*. (Cited in Aschauer, 1990. Rosenstein-Rodan and Rostow-type arguments.)

Musgrave, Richard A. 1959. *The Theory of Public Finance*. New York: McGraw-Hill.

—. ed. 1965. *Essays in Fiscal Federalism*. Washington, D.C.: Brookings Institution.

MVA Consultancy, London; Institute for Transport Studies, Leeds; Transport Studies Unit, Oxford. 1987. *The Value of Travel Time Savings*. Newbury, Berks: Policy Journals. (Cited approvingly in Quarmby, 1989.)

National Academy of Science-National Research Council, U.S.A. 1960. *Conference on Transportation Research*. NAS-NRC publication 840. Washington, D.C.: NAS-NRC. (Cited in Kraft, 1971 as a useful survey.)

- . 1961a. *Transportation Design Considerations*. NAS-NRC publication 841. Washington, D.C.: NAS-NRC. (Cited in Kraft, 1971 as a useful survey.)
- . 1961b. *U.S. Transportation Resources, Performance, and Problems*. NAS-NRC publication 841-S. Washington, D.C.: NAS-NRC.
- National Council on Public Works Improvement. 1988. *Fragile Foundations: A Report on America's Public Works*. (Final report to the President and Congress.)
- Nau, Henry R. 1990. *The Myth of America's Decline: Leading the World Economy Into the 1990s*. New York: Oxford.
- Norrie, Kenneth H. and Michael B. Percy. 1988. *Services and Regional Economic Performance: An Exploratory Analysis*. Halifax: Institute for Research on Public Policy.
- Nye, Joseph S. 1990. *Bound to Lead: The Changing Nature of American Power*. New York: Basic Books.
- Olson, Mancur. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation and Social Rigidities*. New Haven: Yale University Press.
- Ostry, Sylvia. 1990. *Governments and Corporations in a Shrinking World: Trade and Innovation Policies in the United States and Japan*. New York: Council on Foreign Relations.
- Ott, Attiat F. and Thomas D. Austin. 1980. "Capital Formation by Government." In *The Government and Capital Formation*. Edited by George M. von Furstenberg. Ballinger, Mass.
- Oum, Tae H. and David W. Gillen. 1983. "The Structure of Intercity Travel Demands in Canada: Theory Tests and Empirical Results," *Transportation Research*, 17B, 3.
- Oum, T. H. and M. W. Tretheway. 1988. "Airline Hub and Spoke Systems," *UBC Commerce W.P.* 1274.

- Oum, Tae, W. G. Waters II and J. S. Yong. 1990. "A Survey of Recent Estimates of Price Elasticities of Demand for Transport." (World Bank working paper.)
- Oum, T. H. and Y. Zhang. 1990. "Transportation Infrastructure Pricing in the Presence of Lumpy Investment." Discussion Paper No. 16. Kingston: Queen's University, John Deutsch Institute for the Study of Economic Policy.
- Owen, Wilfred. 1957. *The New Highways: Challenge to the Metropolitan Region*. Hartford, Connecticut: Connecticut General Life Insurance.
- Palmer, John P. 1988. *An Economic Analysis of Canada's Ground Transportation Sector*. In *The Economics of the Service Sector in Canada*. Vancouver: The Fraser Institute.
- Peterson, George E. June 1990. "Is Infrastructure Being Under-Supplied?" In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell. (Included are important comments on this paper by Alan S. Blinder and Joel A. Tarr.)
- Pigou, A. C. 1918. *The Economics of Welfare*. London: The Macmillan Company.
- Porter, Michael E. 1990. *The Competitive Advantage of Nations*. New York: The Free Press.
- Quarmby, D. A. 1989. "Developments in the Retail Market and their Effects on Freight Distribution" *Journal of Transport Economics and Policy* 23, 1.
- Ranis, G and J. C. H. Fei. 1964. *Development of the Labour Surplus Economy: Theory and Policy*. Homewood, Ill. Irwin.
- Reuber, Grant L. 1969. "Discussion: Professor Arrow on the Social Discount Rate." In *Cost-Benefit Analysis of Manpower Policies*. Edited by G. D. Somers and W. D. Wood. Queen's University, Industrial Relations Centre.

- Romer, Paul M. 1989. "Capital Accumulation in the Theory of Long-Run Growth." In *Modern Business Cycle Theory*. Edited by Robert J. Barro. Cambridge, Mass.: Harvard University Press.
- Rosenstein-Rodan, P. N. 1961. "Notes on the Theory of the Big Push." In *Economic Development for Latin America*. Edited by Howard S. Ellis and Henry C. Wallich. New York: St. Martin's Press.
- Rostow, Walt W. 1960. *The Stages of Economic Growth: A Non-Communist Manifest*. Cambridge, U.K.: Cambridge University Press.
- . 1962. *The Stages of Economic Growth*. Cambridge, U.K.: Cambridge University Press.
- . 1987. "Reflections on the Drive to Technological Maturity." *Banca Nazionale Del Lavoro Quarterly Review*, 161.
- Royal Commission on National Passenger Transportation. 1991. *Getting There: The Interim Report of the Royal Commission on National Passenger Transportation*. Ottawa: Supply and Services Canada.
- Salvas-Bronsard, L. and E. Bastien. 1984. "A Note on the Estimation of Complete Demand Systems from Canadian Household Budget Data." *Canadian Journal of Economics* 17, 1.
- Savoie, D. J. 1981. *Federal-Provincial Collaboration: The Canada-New Brunswick General Development Agreement*. Montreal: McGill-Queen's University Press.
- . 1986. *Regional Economic Development: Canada's Search for Solutions*. Toronto: University of Toronto Press.
- Scarfe, Brian and David Krantz. 1988. *The Market for Hospitality: An Economic Analysis of the Accommodation, Food and Beverage Industries*. Vancouver: Fraser Institute.

- Schultz, Richard J. and Alan Alexandroff. 1985. *Economic Regulation and the Federal System*. Volume 42 of the research studies of the Royal Commission on the Economic Union and Development Prospects for Canada (the Macdonald Commission). Toronto: University of Toronto Press.
- Schultze, Charles L. 1974. "Sorting Out the Social Grant Programs: An Economist's Criteria." *The American Economic Review*, May 1974.
- . 1990. "The Federal Budget and the Nation's Economic Health." *Setting National Priorities: Policy for the Nineties*. Washington: The Brookings Institution.
- Scott, A. D. "A Policy for Declining Regions: A Theoretical Approach." In *Regional Economic Policy: The Canadian Experience*. Edited by N. H. Lithwick. Toronto: McGraw-Hill Ryerson.
- Simeon, Richard and Ian Robinson. 1990. *State, Society, and the Development of Canadian Federalism*. Vol. 71 of the research studies of the Royal Commission on the Economic Union and Development Prospects for Canada (the Macdonald Commission). Toronto: University of Toronto Press.
- Sitwell, O. F. G. and N. R. M. Seifried. 1984. *The Regional Structure of the Canadian Economy*. Toronto: Methuen.
- Slater, David W. et al. 1991. *The Contribution of Investment and Savings to Productivity and Economic Growth in Canada*. Working Paper No. 10. Ottawa: Investment Canada. (Forthcoming March 1992.)
- Small, Kenneth A., Clifford Winston and Carol A. Evans. 1989. *Road Work: A New Highway Pricing and Investment Policy*. Washington, D.C.: The Brookings Institution.
- Solow, Robert. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, August 1957.
- . 1962. "Technical Progress, Capital Formation and Economic Growth." *American Economic Review*.

- . et al. 1989. *Made in America*. Cambridge, Mass: The MIT Press.
- Statistics Canada. *National Income and Expenditure Accounts*. (Various issues.)
- Stuber, Gerald. 1986. *The Slowdown in Productivity Growth in the 1975–83 Period: A Survey*. Bank of Canada Technical Report No. 43.
- Studnicki-Gizbert, Konrad W. 1990. "Transport Subsidies." (Chapter of unpublished book made available by author to the Royal Commission.)
- Sullivan, Edward C. et al. January 1989. *Transportation and Economic Development of Coastal Regions in the Pacific Northwest*. Newport, Oregon: National Coastal Resources Research and Development Institute.
- Swimmer, David. 1989. *Perspective on Canadian Investment*. Ottawa: Investment Canada.
- . 1990. "Infrastructure." Paper prepared for Investment Canada; unpublished.
- Transport Canada. 1978. *Transportation: A National and Regional Perspective*.
- . 1980. *Passenger and Goods Transportation: An International Comparison, 1960–1977*.
- . 1982. *Multimodal Passenger Terminals: A Canadian Analysis*. TP 3338.
- . 1988a. *Freedom to Move in a New Transportation Environment*.
- . 1988b. *Aviation Forecasts 1988-2001: Aviation 88*.
- . *National Highway Policy for Canada*. 1988. *Report re Phase 1*.
- . *National Highway Policy for Canada*. 1989. *Report re Phase 2*.
- Transport Research Board, U.S. National Research Council. 1989. *A Look Ahead to Year 2020*. Washington, D.C.

- U.K., Department of Transport, 1987. *Values for Journey Time Savings and Accident Prevention*. (Reference in Quarmby, 1989.)
- Universities-National Bureau Committee for Economic Research. 1965. *Transportation Economics*. New York: National Bureau of Economic Research. (Cited in Kraft, 1971 as a useful survey.)
- Urquhart, M. C. 1988. "Canadian Economic Growth, 1870-1980." Discussion Paper No. 734. Kingston, Ontario: Queen's University.
- Usher, Dan. 1982. "The Private Cost of Public Funds: Variations on Themes by Browning, Atkinson and Stern." Discussion Paper No. 481. Queen's University Institute for Economic Policy.
- U.S. Congressional Budget Office. 1983. *Public Works Infrastructure: Policy Considerations for the 1990s*. Washington, D.C.
- . June 1986. *Federal Policies for Infrastructure Management*. (Cited in Hickling, December 1990.)
- . September 1988. *New Directions for the Nation's Public Works*. (Cited in Hickling, December 1990.)
- U.S. Department of Transportation. 1989. *Moving America: A Look Ahead to the 21st Century*. Proceedings of the conference conducted by the Transportation Research Board and sponsored by the U.S. Department of Transportation: July 24, 1989, National Academy of Sciences, Washington, D.C.
- U.S. Department of Transportation, Transportation Systems Center. 1983. *Highways and the Economy*. Cambridge, Mass. (Cited in Choate, 1984.)
- U.S. Governors' Task Force Report. 1989. Presented to National Governors Association, Chicago. Ill. (Reference in Winston, 1991.)
- Vernon, R. 1969. *Metropolis*, 1985. Cambridge, Mass.: Harvard University Press.

—. 1966. "International Investment and International Trade in the Product Cycle." *Quarterly Journal of Economics* 80.

VIA Rail. 1989. *Review of Passenger Rail Transportation in Canada*.

Watson, William G. 1988. *National Pastimes*. In *The Economics of the Service Sector in Canada*. Vancouver: The Fraser Institute.

Winston, Clifford. 1985. "Conceptual Developments in the Economics of Transportation: An Interpretive Survey." *Journal of Economic Literature* 23. (Cited in Winston, 1991.)

— 1990. "How Efficient is Current Infrastructure Spending and Pricing?" In *Is There a Shortfall in Public Capital Investment?* Edited by Munnell. (Also included are important comments by Alan A. Altshuler and Michael E. Bell.)

—. 1991. "Efficient Transportation Infrastructure Policy." *Journal of Economic Perspectives* 5, 1.

Woodward, Robert S. 1974. "The Capital Bias of DREE Incentives." *Canadian Journal of Economics* 7, 2.

SUBSIDIES IN CANADIAN PASSENGER TRANSPORTATION

Trevor D. Heaver

July 1991

I. INTRODUCTION

The Royal Commission on National Passenger Transportation was established in October 1989. Its mandate is to make recommendations for an integrated national passenger transportation system that will serve Canadians well into the 21st century.

The Commission provides an opportunity for a much needed review of a wide range of practices and policies that affect the characteristics and costs of passenger transportation services in Canada. This paper is a part of the background research of the Commission. Its purposes are to provide, in a concise readable form, a comprehensive survey of the nature and role of subsidies affecting passenger transportation services in Canada.

The Commission has a unique opportunity to examine the passenger transportation system of Canada as a whole. It is important, therefore, that research which reports on the historical record of passenger services in Canada look beyond the modal orientation of past practices. This paper considers the general characteristics of subsidies, examines the attributes of subsidies in the individual modes and, finally, draws general conclusions.

OBJECTIVES OF THE PAPER

The terms of reference for the paper encompass the following objectives:

- identification of the linkages, if any, between the types of subsidy, the mechanisms for the provision of subsidies, and the institutional characteristics of the services involved;
- assessment of the effectiveness of subsidy types and mechanisms; and
- assessment of the potential role of subsidies in the future.

OUTLINE OF THE PAPER

The paper is divided into four sections. Following the Introduction, Section II describes the various characteristics of subsidies, including the types and purposes of subsidies, the economic effects of subsidies and the desirable attributes of subsidy mechanisms. Indeed, there is such a variety of subsidies that the second section of the paper begins with a definition of "subsidy."

The frameworks used to describe subsidies provide structures for the survey of the role of subsidies in the various modes of passenger transportation in Canada. Policies and programs exist on a modal basis for a variety of reasons. For convenience, the survey is presented by mode in Section III, concluding with a synopsis of the record of subsidies.

Section IV presents conclusions based on the Canadian experience to date which provide a basis for recommendations on the direction of policy and on further research to assist the framing and application of this policy.

II. CHARACTERISTICS OF SUBSIDIES

This section provides an overview of the nature of subsidies affecting passenger transportation services in Canada. Because of the great diversity of specific subsidies, they have been placed into broad categories. At times, this has involved making some somewhat arbitrary allocations.

DEFINITION OF "SUBSIDY"

The Commission is considering the impact of a variety of government programs and policies on the roles of the various modes of transport in serving the passenger transportation requirement of Canadians. This paper deals with those programs and policies that involve subsidies.

The breadth of perspective adopted here is appropriate to the broad mandate of the Commission. The Commission needs a perspective on the influence of subsidies in passenger transportation equivalent to that being developed by the ministers of agriculture in various countries who are trying to develop a common international measure of the effect of subsidies on agriculture. Their solution, the Producers Subsidy Equivalent Measure, places a wide range of policies into a consistent framework which allows the effect of diverse programs on agriculture to be measured. Faced with a wide range of subsidies in passenger transportation, the Commission has a similar task.

The term "subsidy" carries a variety of meanings. Subsidies are most frequently thought of as grants or payments from the government to aid a particular group. This interpretation is the basis of many studies of subsidies. Subsidies, however, may take a variety of forms. They may be implicit as well as explicit. They may also be accidental as well as deliberate. They may result in the absorption of private costs by the public rather than an actual transfer of public resources to private interests. All subsidies, however, increase the value of benefits relative to costs accruing to the providers or users of a service. Therefore, they affect the output of individual services and the relationship between competing services. They affect the role of the modes of transportation in serving Canadians' needs.

For this paper, subsidy is defined as "a transfer of benefits to or costs from an affected party by an implicit or explicit program or policy." This definition is sufficiently broad to encompass the wide variety of types of subsidy considered below. The result of such programs or policies is usually a greater output of the subsidized good or service than warranted in the absence of the subsidy.

The immediate expense of a subsidy is a cost to society, whether it falls on some or all members or whether its effects are channelled through government. A subsidy program should only be continued willingly and

knowingly, if the total value of benefits exceeds the total value of costs. A subsidy might be made available to all industries — for example, a grant in relation to research expenditures — or it might be made available selectively. A subsidy will often involve different treatment of one firm or industry from another. For example, the management of international shipping has been given tax-free status in Canada under particular conditions, effective March 1, 1991. This is classified as a subsidy here. The effects of the subsidy and whether it has merit are separate issues to be considered in the light of domestic and international conditions of this industry.

TYPES OF SUBSIDY

Several means exist by which income may be transferred with the result that the output of a good or service is larger than would otherwise be the case. There are four categories of such means, each requiring an explicit or implicit public decision. The public may pay for services; the public may provide services; the public may protect services; or the public may absorb costs created by services. Exhibit 1 is a summary of the types of subsidy discussed here.

This classification structure encompasses a broader range of subsidies than are normally considered. Most studies focus on direct subsidies involving payments; these are most easily identified. However, the full effect of government intervention requires a broader approach. A report by the Organisation for Economic Co-operation and Development on government intervention places subsidies in six specific classes.¹ They are cash expenditures; preferential credits; tax expenditures; subsidy equivalents of regulatory measures; possible subsidy elements in public purchasing contracts; and subsidy equivalents of tariffs and non-tariff barriers to trade. The classification used here places subsidies in fewer, generic classes. It has the advantages of highlighting the general nature of the promotional policies and of avoiding detailed technical listings.

Exhibit 1

SUMMARY OF TYPES OF TRANSPORT SUBSIDY

A. Subsidies which the public pays

- *ex post* compensation for unremunerative services
- *ex ante* compensation
- support for research and development
- tax concessions

B. Subsidies when the public provides facilities or services

- unremunerative services provided to travellers
- operation of unremunerative facilities and related services
 - routes
 - terminals
 - vehicles (not applicable to passenger transport)
- implicit capital subsidy of Crown corporations

C. Subsidies associated with the protection of services

- cross subsidies by regulated monopolies

D. Subsidies by the absorption of costs

- loan guarantees
- environmental externalities

Subsidies Which the Public Pays

Subsidies which the public pays may come in a variety of forms. Some may involve actual payments, others may enhance the net cash flow of undertakings by providing tax concessions. The programs are described in four categories: *ex post* compensation, *ex ante* compensation, payments for research and development and tax concessions.

Ex post compensation: Payment to a company for an imposed public service provided at a loss, but which is required in the interests of society, is the most obvious form of a subsidy in transport. It is compensation to the provider of the service for the losses incurred. The primary beneficiaries of the subsidy are the users of the service. After the *National Transportation Act* (NTA, 1967) and until the formation of VIA Rail in 1978, CN and CP received compensation for 80 percent of the losses they incurred in the provision of intercity passenger services. This was an *ex post* payment system.

Ex ante compensation: The *National Transportation Act, 1987* (NTA, 1987) has provision for an *ex ante* subsidy system should it be desirable to maintain airline services existing in 1987 but which are no longer remunerative. Section 85 of the NTA, 1987 requires that unremunerative services that are to be maintained should be contracted following competitive tendering.

The payment of compensation to carriers for losses resulting from the imposition of a public duty was an important recommendation of the MacPherson Royal Commission in 1961. It was incorporated in the statement of national transportation policy in the NTA of 1967 and is unchanged in the NTA, 1987. Paragraph 3.(1)(f) of the latter Act states: "each carrier or mode of transportation, so far as practicable, receives fair and reasonable compensation for the resources, facilities and services that it is required to provide as an imposed public duty."

Compensation is appropriate for two reasons. First, as transportation markets become more competitive, it is no longer possible for carriers to earn a sufficiently high profit in some markets to offset losses in another. The effect of competition on the railways made this abundantly clear to the MacPherson Commission. Second, even if the cross subsidy of one service by another is possible, it is inappropriate for one sector of society, identified by chance through the importance of a particular service to them, to pay the subsidy decided appropriate by society at large. This would amount to selective taxation.

Using the railways as an example, and supposing that cross subsidies were possible, it would most likely be the shipper of bulk natural resources from remote locations that would pay the high rates to cover the cross subsidy requirement. A cross subsidy is possible when an enterprise is able to earn some economic rents or excessive profits in a market with limited competition. An inefficiently small output is produced in the "taxed" market.

Payments for research and development: Subsidy payments may be made to carriers for the support of research and development otherwise unattractive to a company. No actual service to the public may be provided immediately. However, on occasion, a service may be involved, as with the experimental short-take-off-and-landing (STOL) service operated by Air

Canada between the island airport in Montreal and Ottawa between 1973 and 1976. Users will benefit coincidentally from the greater range of service available (while some other suppliers may be losers).

Tax concessions: Various forms of tax concessions may be used to enhance the cash flow of tax-paying organizations and, thereby, stimulate certain activities. Examples of concessions could include exemption from consumption taxes otherwise payable, for example, a fuel tax; tax credits for certain expenses incurred, for example, research expenditures; and acceleration of capital cost allowances on particular classes of assets for an industry.

Tax concessions may be viewed as “tax expenditures,” a term used to denote the tax revenue foregone. Subsidies provided by tax concessions are less visible than direct payments. They are also less readily measured. The amount of tax foregone on a consumption tax may be easier to estimate than the amount foregone through accelerated capital cost allowance. The latter only applies when firms are earning profits.

Subsidies When the Public Provides Facilities or Services

Subsidies when the public provides facilities or services arise in a variety of situations when the revenue generated by the facilities or services do not cover their costs. The immediate beneficiaries are the users of those facilities and services. If these beneficiaries are suppliers of services in a competitive market they may be forced to pass on the benefits of the subsidy to consumers. The subsidies that exist in the provision of facilities and services to carriers are in spite of the principle advanced by the MacPherson Commission and reflected in the NTA, 1987, paragraph 3.(1)(e), namely, that “each carrier or mode of transportation, so far as practicable, bears a fair proportion of the real costs of the resources, facilities and services provided to that carrier or mode of transportation at public expense.”

Unremunerative services provided to travellers: The public provides some services directly to the public without covering the costs from user revenue. Ferry services are an example. They range from the extensive deep-water services in Atlantic Canada to short crossings of rivers throughout Canada. VIA Rail is also an obvious example, although it is appropriate to note that VIA purchases a significant component of the transport operation from CN and CP Rail.

Unremunerative facilities and services for routes: The public also provides infrastructure and services which must be used in conjunction with other assets to create transportation services. Public expenditure provides routes both in the form of physical structures and of services. The largest public investment in the nation is the network of roads provided for the operation of private and for-hire vehicles. As well, navigation aids and a variety of services are provided to facilitate the operation and safety of air and marine services.

Unremunerative facilities and services for terminals: Passenger terminals are currently provided at public expense in the case of airports, and, in the special case of cruise ships, in ports. For both routes and terminals, the losses incurred may be visible in the sense of shortfalls in revenues against costs, although they may be difficult to quantify accurately. Hidden subsidies may also exist — for example, the absence of taxes on public land. This is a subsidy in the sense that competing land uses have to pay taxes.

Implicit capital subsidies of Crown corporations: The public provision of services by government may involve subsidies which are hidden in the capital structure. The cost of capital to a government agency is reduced by the assurance provided by state backing. The cost of capital does not reflect the true risk associated with the investment undertaken. The effect of risky projects on the cost of capital is a continuing concern for private firms, even though the effects of independent projects are averaged across the activities of a firm.

Subsidies Associated With the Protection of Services

The output of firms may be made more profitable by providing protection against competition. This may be achieved by tariffs or quotas that keep out foreign goods, thus protecting domestic goods or services by the regulation of entry into the industry.

The transportation industry has been subject to a wide range of economic regulation, including entry control for a variety of reasons. They include the avoidance of destructive competition, protection of “infant industries,” and the interest in preserving the ability of carriers to meet public service or common carrier obligations. The latter category has really meant that carriers sustain unremunerative services by cross subsidizing unprofitable services from profitable ones. The cross-subsidy structure has been

sustained by precluding competition on the profitable routes. The effect is to support a subsidy mechanism through selective taxation achieved through the regulatory process. A recent high-profile example of this policy is the regulation of entry into the telephone industry and the cross subsidization between long distance and local calling services in that industry.

An example of this policy in passenger transportation is in the regulation of bus services in Quebec and Saskatchewan, where there is strong concern to preserve services to small communities. The latter concern had also been reflected in the service mix operated by Air Canada when it was a Crown corporation.

Regulation may be introduced and administered to achieve cross subsidization. However, it may also simply create an environment in which cross subsidization evolves because of the passive role of authorities in regulating rates and because of the diminished market discipline faced by carriers.

Cross subsidies most often become evident when firms are faced with a loss of earnings from their protected markets. For example, cross subsidies have become issues as competition from other modes erode the earnings of main bus routes and as potential competition stimulates "rebalancing" in telephone rates.

Subsidies By the Absorption of Costs

The costs of doing business may be absorbed by society in two quite different ways. The first is frequently recognized as a form of indirect subsidy. The second is less often thought of as a subsidy.

Loan guarantees: Loan guarantees are frequently used by government to enable capital to be available to companies, or available to companies at lower rates than otherwise would apply. The result is that the costs associated with the possibility of the venture failing are shifted to government.

Environmental externalities: Externalities are the consequences, positive or negative, of an activity on others and for which the supplier does not receive or pay compensation. Externalities have long been recognized as important in transport. For example, the consequences of added traffic on a route, whether highway, airline or canal, can be delay to other vehicles. The immediate effects of congestion are internal to the particular transport

system. They result in inefficiency but not directly to the subsidization of the mode. Air and noise pollution, however, are externalities exacerbated by congestion; they are examples of costs imposed on society. These externalities fall on society in a general sense, not on government. The incidence of externalities within society varies with their specific nature.

Differences in accident rates among modes of transportation are also associated with an externality to the extent that the health care costs resulting from accidents are borne by government and not by the affected travellers.

The environmental consequences of transportation are becoming more widely recognized. While the costs are being reduced by a variety of measures, such as non-leaded gasoline and mandatory seat belts, they are not eliminated. Their values are difficult to estimate but differ among the modes. The effects of their absorption by society on modal output is uncertain. (It should be noted here that the environmental performance of modes must be judged on the basis of their effects in relation to actual passenger transportation performed and not on the basis of their potential performance.)

PURPOSES FOR SUBSIDIES

In view of the various types of subsidies, it is not surprising that while all subsidies have explanations, some have more explicit purposes than others. Before these purposes are discussed, it is helpful to recognize important attributes of subsidy programs in practice.

First, not all subsidies are the result of deliberate programs. While many subsidies are intentional, others are coincidental and still others are accidental. An example of a deliberate subsidy is payment to VIA, but the saving in the cost of capital attributable to Crown corporations is coincidental. Such a coincidental subsidy could reasonably have been anticipated, but the subsidy flows from the program of Crown ownership which was introduced for other reasons.

Some subsidies have come into existence "by accident." An example of such an accidental subsidy is in the failure of the St. Lawrence Seaway to come up to early expectation and cover its costs. This situation produces a

“subsidy” in the sense that the users will not be able to meet the system costs. However, where a loss arises from an investment decision which, with the benefit of hindsight, was inefficient, no “subsidy” should be considered to exist. The sunk costs are now irrelevant. In the long run, there will be questions about the ability of the service to cover future costs.

Second, subsidy programs frequently lack clarity of purpose and recognition of costs. Given the variety of political considerations surrounding many subsidies, obfuscation may be politically expedient. However, as will be argued later, it can be a major contributor to the high cost and ineffectiveness of a subsidy in achieving desired objectives.

Subsidies may be introduced to achieve a variety of purposes. This is particularly true of the transportation industry which affects so many aspects of society. Although individual subsidies may contribute to several purposes related to various broad goals of society, the reasons for subsidies to passenger transportation are restricted to three broad categories: contributions to nationhood, change in income distribution and increase in economic efficiency; these are listed below in Exhibit 2.

Exhibit 2

SUMMARY OF THE PURPOSES FOR SUBSIDIES

A. Contributions to nationhood

- defence, justice and social services
- independence
- political unity
- prestige
- preservation of culture

B. Income redistribution

C. Economic efficiency

- public goods and services
- externalities
- decreasing costs
- inadequate information

Contributions to Nationhood

Transportation contributes to nationhood in various ways. The relative importance of these contributions changes over time and varies between countries, but collectively they exercise continuous and strong influences.

Defence, justice and social services: An independent, sovereign state must have the mobility as well as the resources to ensure the administration of justice and the performance of national defence. Historically instrumental to the construction of early trunk roads, these fundamental requirements still have their relevance. The establishment and maintenance of Canadian sovereignty in the Arctic islands require a sufficient Canadian presence — a presence which still must be supported from government budgets.

In the social area, a current transportation-dependent service which Canadians might expect to be able to provide to all communities is access to health care. While not a necessary attribute for a nation, it may be a fundamental service expected for communities, irrespective of their economic opportunities. Such accessibility is provided by air transport.

Independence: Transportation played well-known roles in ensuring Canadian independence and sovereignty. Each mode of transportation has its example. Observation of Canadians' behaviour in selecting their travel routes today suggests that the need to preserve independence by "buying Canadian" no longer seems relevant. However, in political debates, the preservation of Canadian alternatives, be they companies or routes, is a well-known argument. How important independence is in this argument and what it currently means are not clear. It is certainly in the minds of some. However, the protection of their own capital and of their own jobs may be in the minds of others.

Political unity: The basis for political unification of disparate geographical units into one Canada has been achieved, in part, by offsetting the economic cost of distance with subsidies. Railway service and, later, the reduced freight rates for Atlantic Canada, and the promise of a railway to British Columbia were the early forms of this. The later extension of the Maritime freight subsidy to trucking in Atlantic Canada, and the operation of buses for rail passenger service in Newfoundland have been efforts to make the subsidies more cost effective.

Prestige: National symbols are important. Sports heroes such as Pelé and Gretzky, and participation in the Olympics are evidence of this. So too, apparently, are national airlines, given the number of countries that have one. While Canada, unlike many developing countries, has resisted the need for a national shipping line, we are now testing the importance which we attach to Canadian-owned airlines. (Related issues in the current airline policy debate include the competitiveness of markets and the equality of opportunity for Canadian and foreign firms to behave and compete in comparable ways.)

Preservation of culture: Assertion of community cultural values is a strong phenomenon of our times. Within Canada, it may take the form not only of greater identity within various ethnic and cultural groups, but also of sustained concern for preserving Canadian values in the face of the Free Trade Agreement and broader trends to globalization.

The environment gives rise to concerns for the preservation of “things Canadian,” quite possibly including the continuation of transcontinental rail passenger service. The argument might go: “Canada would not be the same without it; the railway laid the foundation of the country.” If unremunerative rail service is to be retained, the Commission may have to decide whether rail service has cultural value in an evolving and dynamic society. Of course, there may be other reasons for subsidizing rail service. (This does not deny that there may be economic as well as cultural value in operating museums, as the number of steam-engine rail services and paddle-wheel cruises testify.)

Income Redistribution

Regionalism is strong in Canada for geographic and cultural reasons. The country early used railways as an instrument to attract remote regions into Confederation. Subsequently, the subsidization of transportation was used to promote regional economic development. In effect, this is a way of taking wealth from the “haves” and using it to enhance the economic opportunity of the “have nots.” The Hudson’s Bay railway is a costly example.

Transportation may also be subsidized to provide enhanced mobility for particular sectors of society. For example, it is argued that the maintenance of subsidized rail service facilitates the mobility of poor and physically

disabled people. (Rail service can offer persons with disabilities better mobility while travelling than can bus service.) The requirement that carriers provide extra facilities for disabled persons could result in services at less than cost. In the absence of public compensation, this would require cross subsidies.

Economic Efficiency

There are many arguments for the provision of subsidies to increase the efficiency with which a nation's resources are used. Subsidies may be used to enhance the efficiency of the economy rather than to redistribute wealth for public reasons but at an economic cost. However, subsidies introduced for efficiency reasons will have some distributional consequences. They may also have some offsetting side effects on efficiency in the economy, as discussed in the next subsection. Subsidies intended to enhance efficiency have their costs as well as benefits.

The economic arguments for such subsidies can be placed in four categories. They are: inefficiencies in the provision of public goods and services; the influence of externalities; decreasing costs; and inadequate information.

Public goods and services: Public goods and services cannot be efficiently marketed. A consumer can derive a benefit from the service without adversely affecting consumption by others. The classic example is the service provided by a lighthouse which is available to all ships without any effect on lighthouse costs. The marginal cost is zero; therefore, the appropriate price is zero.

The view of roads as public goods was important during the early development of roads when they primarily provided property access and were, in general, little used. The increased use of roads and their use for long-distance travel rather than access, have changed their nature. The costs of providing additional road capacity for users have become significant, and control of access to divided highways through tolls or similar means has become practical.

Externalities: Economists' models of perfect competition assume that suppliers and consumers experience the full consequence of their actions. This is not the case in reality. For example, transportation gives rise to a

number of effects felt beyond the individual supplier or user. Externalities, as noted earlier, are relevant here as they may give rise to conditions warranting subsidies.

Where externalities are positive, a subsidy may be appropriate to encourage an expansion of output to an optimum level. A frequently cited example is the case for a subsidy to orchard owners to reflect their contribution to the value of honey produced by separate beekeepers. An example may occur in transport when an increase in the number of users of a scheduled service enables service frequency to be increased, to the benefit of all users. The result for a service charging marginal cost, taking into account producer and user costs consistent with economic theory, is that marginal cost is declining and that a subsidy is necessary for the service to be financially viable.

Transport services produce a number of negative externalities. Most widely recognized are air and noise pollution. These externalities may be dealt with directly through taxes or regulations, or by encouraging other activities or modes of transport which do not produce these negative externalities. Subsidies are suggested for bus and train to offset the negative externalities of the automobile.

This is the appropriate place to discuss the issue of energy conservation. There are four points to make. First, if energy prices reflect scarcity, there is no reason for specific intervention. Second, energy is only one of a number of inputs. Forced conservation of energy will likely result in waste in other resources such as capital. Third, energy conservation at one stage in production may result in higher energy consumption at others. It is necessary to look beyond the immediate process. Fourth, energy consumption must be related to the work performed, not the technological possibilities. For example, a full aircraft may require less oil per passenger trip than a train with 20 percent occupancy.

Decreasing costs: Charging an efficient price equal to marginal cost will result in a loss when a firm's cost per passenger or ton falls as the total amount of traffic carried increases. This condition may arise when there are increasing returns to scale — for example, larger vehicles are inherently more efficient — or if there is a surplus of capacity which may be caused (temporarily) by lumpy investment. Losses may be avoided in two ways. One method is for the public to subsidize the service. An alternative method

is to allow the carrier to practice price discrimination, that is, to charge different users different prices. This is the approach used by airlines in seat management and by VIA in the pricing of services.

A particular form of the increasing-return argument is for infant industries which may not currently be efficient or viable because of small size but subsequently should grow into viable and profitable firms. The infant industry may justify “start-up” subsidies; the early protection of Trans-Canada Airlines might have fallen under this category.

Inadequate information: A final condition that may justify a subsidized activity is inadequate market information. For example, consumers may lack knowledge which may justify government support of consumer groups.

THE ECONOMIC EFFECTS OF SUBSIDIES

Whatever the reasons for subsidies, it is important that their effects be continually assessed. From the time that they are inaugurated, it is appropriate to consider the benefits and the costs, and how the effects vary with the level of subsidy. Questions should be asked not only about the existence of a subsidy but, if justified, whether it is at the right level.

The economic effects of subsidies are derived from the obvious: subsidies change behaviour. The major difficulty with the analysis of subsidies is the impracticality of forecasting all of the effects *ex ante* or of identifying and measuring them *ex post*. In a dynamic world this is of little surprise, but it is of no comfort.

A study of government intervention in OECD countries² notes that it is impractical to carry out theoretically correct comprehensive analyses of subsidies. The absence of detailed data requires levels of assumption that makes partial analysis more appropriate.

The effects of subsidies depend on attributes of the buyers and sellers of goods and services. These attributes are discussed in the following pages before an examination of the nature of the economic costs of subsidies and of the unintended and, generally, indirect effects of subsidies. The section concludes with a consideration of some issues raised by the incomplete recovery of costs for publicly provided facilities and services.

The Importance of Buyer and Seller Responses

Subsidies are designed to change the consumption of particular goods or services. The extent to which they achieve this is dependent on the elasticity of supply and demand, that is, the extent to which the behaviour of the buyer and seller is influenced by changes in price. If the amount supplied and demanded are both highly responsive to price, a significant change in output will result from a subsidy. If neither are responsive to price, little change in output will be achieved.

Further, the extent to which a subsidy is consumed by suppliers' cost increases rather than realized by buyers in increased consumption is dependent on the relative elasticity of supply and demand. In a situation where suppliers face sharply increasing costs, much of the subsidy will be consumed by the supplier. If supply could be increased without limit at a constant cost, the benefits of a subsidy would accrue to users.

Assessment of the effectiveness and incidence of subsidies is critically dependent on a knowledge of the elasticity of supply and demand for the service concerned. The elasticity of demand is affected by the availability of substitute services. The elasticity of supply is affected, also, by the amount of competition.

When buyers or users of a transportation service benefit from a subsidy, it is important to distinguish between those that realize a "windfall" gain, and those that benefit from the subsidy shift. Travellers that would have been using a service anyway, get the windfall gain. Those that use the service because of the subsidy are the intended targets of the subsidy. (Given the general willingness of foreign tourists to pay a much higher [probably commercial] rate for rail service, it would be interesting to know the proportion of foreign tourists that use VIA service between British Columbia and Alberta!)

The Direct Economic Costs of Subsidies

In theory, the economic cost of a subsidy is not measured by the expenditure of the subsidy, whether in the form of cash payment or tax revenue foregone. Such expenditures are merely transfers from one group in society to another. The economic cost should be measured by the value of output foregone as a result of the transfer. That depends on conditions

in the economy and the mechanism by which the revenue is raised. As a practical matter, payments (as distinct from tax expenditures) are presumed to be economic costs. They may overstate the economic costs; for example, the social cost of labour may be overstated during a period of high unemployment. However, they are more likely to understate the *total* economic cost.

Costs are incurred not only by the need to transfer resources from one sector to another by taxes, but also by shifts in the consumption of substitute services. For example, subsidies to rail service results in transfer of passengers from air, car and bus services. The substitution effect is believed to be greatest on bus service. The effect is to lose the net value of some bus service, together with the cumulative effect that this has on the value of bus service frequency. A transfer of travel will also take place from car to rail, which may be associated with an immediate net gain because of the negative externalities associated with car travel. In the example, the additional loss in value from bus travel might be offset by the reduction in externalities of car travel.

However, perhaps a more important source of cost, than the inefficiencies resulting from marginal shifts in resources from more to less productive uses, is the behavioural response of buyers and sellers. Presumption that the economy is efficient when prices are set on the basis of appropriate concepts may be seriously in error when suppliers are protected from market forces. This is now a matter of increasing concern to the World Bank.³

Three aspects of behavioural response are important. First, because of subsidies, suppliers may be protected from the rigours of the market and become less innovative and efficient. Second, those providing goods or services to suppliers may be able to achieve some economic "rents," that is, returns higher than competitive market conditions would warrant. This might be true for various inputs provided to VIA. Third, consumers may be lulled into the acceptance of service and productivity conditions not otherwise acceptable. This is currently evident in the response of users to government proposals to increase cost recovery in the marine and air modes. Faced with new charges for facilities and services, carriers are demanding greater efficiency in supply. The collective result of these three responses can be a significant detraction from the efficiency of supply. The indirect costs of subsidies can be large.

The Indirect Effects of Subsidies

The behavioural response of suppliers and consumers just outlined are unintended. Other unintended ramifications of subsidies are of concern here.

Subsidies have indirect ripple effects, the magnitude of which are not easily identified with subsidies. The well-known Canadian example is the effect of subsidies on the movement of grain by rail. The effects include a shift in meat processing out of the Prairie region; concentration on grain monoculture in the Prairies with various economic, environmental and social effects; a stultified trucking industry; and a high-cost grain logistics system. These effects of holding down the rates for rail transportation are well known. Getting rid of the inefficiencies is not easy because groups develop vested interests.

Subsidization of passenger transportation can have comparable effects. Take the case of subsidizing rail and bus services in southern Ontario, to offset the externalities of the car. The result? Low-cost transport and the consumption of more transport than is warranted. Social and economic choices are geared to the low apparent cost of transport. Community growth and travel behaviour are more transport-intensive than would otherwise be the case. Long work trips by rail, as well as by car, may become more common, while low-density living will result in reliance on the automobile for local and other trips.

The examples of grain and passenger transportation demonstrate the important interrelationship that exists between land-use patterns and transportation. The effects vary with the subsidy from local to national in scale. The magnitude of the effects depends on the relative importance of transportation in location choice. When transportation subsidies affect locational decisions, they have deep and long-lasting effects.

Subsidies and Cost Recovery

Some subsidies exist in transport because fees collected from users of publicly provided facilities and services may not cover the costs those users impose on the system. This applies in each of the modes of transport and gives rise to issues worthy of special attention.

Unfortunately, the reasons for less than full cost recovery are not generally known. It may be that subsidization has been intended for reasons of nationhood, income distribution or efficiency. However, it may also be that

practices followed when the national infrastructure was being developed have been continued because of the various technical and political difficulties in implementing user charges. Subsidies may have evolved by default.

The role of infrastructure has changed as the economy has matured (and as technology has changed). Infrastructure which was capital-intensive and little used originally required public support, for example, roads and railways. Traffic on roads has increased greatly; in fact the issue has often shifted to one of dealing with congestion.

Second, economic growth and technological change have diminished the importance of specific facilities to nation-building, and alternate pricing has become practical. The volume of traffic is sufficient to pay for facilities, and practical methods exist or are becoming available to implement charges. This is not the place to examine the various issues and mechanisms for implementing user-charges to achieve cost recovery. It is sufficient to note here that Canada has not and is not pursuing these aggressively! Witness the exclusion of recreational boaters on the recent government proposals for user-charges on waterways. Developments in road pricing and private funding for roads are examples of the technological and attitudinal changes affecting highway finance, especially in other countries.

In theory, efficiency is maximized if tax revenues are raised from sources that have least effect on the (assumed) efficient allocation of resources in the economy. Following this principle, taxes would be increased on goods and services with inelastic demands. This argues for general revenue funding, which provides government with the greatest flexibility in raising and allocating tax revenues. These advantages must be weighed against the tendency toward inefficient supply when users and suppliers are not subject to pricing discipline. The combined result is that arguments are arising for the collection of user-charges sufficient to cover facility costs and for these revenues to be placed in dedicated funds.

SUBSIDY MANAGEMENT

The management of subsidies covers many tasks, including deciding whether a subsidy is warranted, selecting the type of subsidy and managing the subsidy program. A summary of management considerations is presented in Exhibit 3.

Exhibit 3

SUMMARY OF SELECTED SUBSIDY MANAGEMENT CONSIDERATIONS

A. Desirable attributes of subsidy programs

- provide direct subsidies
- subsidies should allow market competition
- subsidies should not distort input costs
- subsidies should be transparent
- subsidies programs should be monitored and results publicized

B. Deciding on a subsidy program

C. Financial subsidy administration

- competitive tendering
- operating tenders
- negotiated service contracts
- *ex post* subsidies

Desirable Attributes of Subsidy Programs

The management of subsidies should be guided by generally desirable attributes of subsidy programs. Subsidy programs are considered effective when resources are used efficiently to accomplish the program objectives and when the cost and effectiveness can be monitored over time. Unfortunately, many subsidies do not have these attributes, and the beneficiaries of subsidies would often rather hide the existence of the subsidy, perhaps for fear of losing the benefit. Five program attributes are recommended to ensure that services are provided efficiently and that desired results are achieved.

Provide direct subsidies: Subsidies are most effective if provided as directly as possible to the intended beneficiary. First, the process requires explicit recognition of beneficiaries, a process likely to cause a critical assessment of the subsidy's merit. (For example, the subsidy per passenger given to tourists going on the Algoma Central tour train might cause taxpayers and competing tourist attractions to take exception.) Second, the effectiveness of the subsidy could be enhanced by leaving the beneficiary with an unbiased choice of transportation to be used, rather than by subsidizing

a mode or carrier. If the “pay-the-farmer” policy makes sense for efficient grain transportation, so does “pay-the-traveller” policy for efficient passenger transportation, if a decision to provide a subsidy to this group has been made.

Subsidies should allow market competition: The effectiveness of subsidy expenditures will be enhanced by mechanisms favourable to competition. Maintaining competition reduces the risk of a subsidy inhibiting innovation and efficiency. An example is the use of competitive service tendering as a means to ensure that, even if a specific service is subsidized, competition is effective periodically, that is, at the time of tendering. Subsidies, including service tendering, should be structured to allow intermodal as well as intra-modal competition as far as possible. When subsidies are introduced to benefit the mobility of travellers, the competitive balance in the transportation market and the effects on innovation are affected least by following the “pay-the-traveller” policy.

Subsidies should not distort input costs: Subsidies which distort management decisions on resource inputs should be avoided, unless they are specifically needed to correct otherwise inefficient input costs. They should not be used as a general form of subsidy. For example, providing tax relief through accelerated capital cost allowances, or providing low-cost capital, provides incentives for an organization to use capital more intensively, perhaps in place of labour or fuel. When capital costs are subsidized, it may be necessary to introduce offsetting investment rules to avoid waste, as evident in the controversy which has surrounded the use of capital by the transit systems in Ontario.⁴

Lump sum payments, whether paid after tendering or negotiation with a selected provider, are least likely to distort supply efficiency. The duration and monitoring of contract provisions are important to the sustained efficient performance of suppliers.

Subsidies should be transparent: Subsidies should be visible and their purposes clearly stated. It is unfortunate that many subsidies lack these important attributes.

Visible subsidies involve explicit financial transfers. Subsidies achieved through concessions and protection from competition are difficult to identify and even more difficult to quantify.

The purposes for subsidies should be spelled out with care. Transport is not subsidized for its own intrinsic value. Subsidies may be provided for transport to remote regions, but it is not “the regions” that are subsidized, but the particular travellers and interest groups.

For example, a rail service into a remote area may serve three broad clientele groups. The first are the local residents who may or may not have alternate means of travel and who live in the area for a variety of reasons. The second group is made up of the travellers who might come to such regions for the fishing or hunting and to stay at recreation lodges and facilities. Rail service may be one form of access providing a benefit to these travellers as well as to the lodges they visit. The third group may be thought of as wilderness travellers who use the rail service to get to a remote location, where they are then self-reliant. When the rail service is subsidized, and low rates are provided to everyone, all travellers are subsidized by the taxpayers; is this intended?

The reasons for subsidizing mobility for residents are likely different from the reasons (if any) for subsidizing other travellers. In order to examine alternatives, the explicit reasons for subsidies must be revealed. (An interesting condition sometimes applies in remote regions served by rail, under which existing lodges do not want lower-cost road service provided, since the presence of more vacationers would detract from the “wilderness” assets that they enjoy.)

Subsidy programs should be monitored and results publicized: Careful, explicit rationales for subsidies provide the basis for effective monitoring programs. The more direct and explicit a program, the easier and the more likely is the conduct of program evaluation. The effectiveness of the subsidy in relation to program goals and the incidence of costs and benefits should be estimated and reported.

Experience with monitoring subsidy programs is not good. Programs involving expenditures are reported through records of departmental and agency budgets and records. However, except where an agency, such as the National Transportation Agency, has responsibility for a subsidy program, reporting is subject to limited visibility. Critical assessment of the effectiveness and incidence of subsidy effects is usually not carried out.

Subsidy program assessment raises difficult questions about whether assessments should be carried out by program officials or by staff with special expertise in evaluation methods. Difficulties also surround the way results are reported. Recipients and their elected representatives are normally reluctant to see subsidies analyzed and the results publicized. Their perspective is commonly dominant, to the detriment of taxpayers in general and, often, to the detriment of subsidized interests themselves as the effectiveness of subsidies diminishes. Subsidies may even become counter-productive.

Deciding On a Subsidy Program

The management of subsidy programs is about choices. For choices to be made well and for programs to be managed efficiently, transparency and explicit decision processes are essential.

To decide whether or not a transportation subsidy is warranted cannot be a technical process. Judgement will always be important in weighing intangibles and making trade-offs between heterogeneous considerations. However, guidelines can help in the selection of effective programs. They are intended to ensure sufficient information to support rational and effective choices, and include:

- Clear identification of program goals, whether related to nationhood, income redistribution or economic efficiency;
- Consideration of transport and non-transport alternatives;
- Consideration of alternate amounts of subsidy;
- Evaluation of alternatives by such techniques as cost-benefit analysis or cost-effectiveness analysis; and
- Detailed description of the expected distribution of costs and benefits.

The guidelines reflect several important aspects of program choice, emphasizing the different means of achieving goals. There is a right size as well as type of subsidy. A knowledge of the distributional effects of a subsidy is different from and must be weighed with other costs and benefits.

Allowances must always be made for uncertainty in subsidy effects, especially in unintended side effects. Directness of subsidies, the preservation

of competition as far as possible, and good monitoring programs help to increase the effectiveness of programs.

Financial Subsidy Administration

Various methods used to provide direct financial support to transportation firms reflect the structure of the industry and government policies. For convenience, four categories of methods are used.

Competitive tendering: Competitive tendering has become more popular for the provision of subsidized services, in conjunction with the move to private provision of public services, as is evident in the United Kingdom, for example. It is most common in Canada in the provision of transit services in small communities. Competitive tendering requires the careful specification of service requirements, itself a beneficial discipline for public (and private) organizations involved in contracting out. The establishment of successful service standards is evident in the contracting of emergency health-care services.

The tendering process is intended to provide the advantages of competition in terms of innovation and efficiency among alternate suppliers. The competitive advantage is realized each time a tender is opened, although the life and versatility of the assets affect the optimum life of the contract period. While an incumbent normally enjoys some advantages on retendering, competition still remains an important influence.

Tendering may allow for some versatility in service design. For example, a community may allow applicants to use large or small vehicles with different service frequency or, alternatively, may define a service level, including the type of equipment to be used. The former approach allows greater scope for innovation.

Services such as bus or air service lend themselves to the tendering process. In rail transportation, the normal service regime provided by the company owning the infrastructure is not consistent with competitive tendering. However, if a separate contract is issued for passenger services to use a railway track, competitive rail tendering would be possible. The practicality of independent rail passenger service is evidenced in Western Canada where one private rail service to Calgary is already approved and a second is being planned.

Operating tenders: While service tendering requires the provider both to supply capital stock and to assume operating responsibility, an operating tender requires only the latter. Capital equipment is provided by the government. This is the case in Ontario, for example, where capital subsidies are used to provide buses, which in some municipalities are operated and maintained by private contractors. While this may result in some advantages derived from centralized buying power and standardized equipment, it has the disadvantage of curtailing innovation and responsiveness to local conditions.

Negotiated service contracts: Where competitive tendering is not possible, negotiated service contracts may be used. Such is the case for BC Rail's passenger services which are subsidized by the province. Negotiated contracts can involve careful service definition and agreed levels of capital investment and operating subsidy. In the case of a "captive" Crown corporation, the quality of the contract is very much dependent on the attitude and ability of the individuals involved. Competition does not provide any check.

The contracts, like tender arrangements, can include penalties for service failures and can provide incentives for the achievement of additional revenues.

Ex post subsidies: Subsidies may be provided to public services on an "as-needed" basis. Under such regimes the level of service may be tightly or loosely defined. In either case, the shortfall will be made up by government. Budgets, however, would normally be subject to government approval. The system operates as a cost-recovery process. Consequently, incentive for innovation may be low but the risks of innovation may also be low, at least for the organization, if not the individuals involved. Overall, the system removes the pressure of bottom-line results and has generally been seen as conducive to waste.

III. THE SUBSIDY EXPERIENCE

The breadth of purpose, types and effects of subsidies outlined in the previous section makes a general description of subsidy programs difficult. Therefore, while certain themes are relevant across the modes of transport, a modal approach is used to review the role of subsidies in Canadian intercity passenger transportation.

HIGHWAY TRANSPORTATION

Since the automobile dominates intercity passenger transportation in Canada, its use must be given careful consideration by the Commission. Relatively small percentage changes in car use can have large absolute values and large potential effects on other modes which are small by comparison. Therefore, the first part of this discussion deals with issues of the provision of highways and of automobile use in general. Characteristics of subsidies involving buses are considered separately.

The discussion does not deal at length with the operation of trucking, although the efficient provision and pricing of roads for passenger transport is unavoidably linked with the efficient use of roads by trucks. The linkage between freight and passenger services must be dealt with by the Commission in each mode of transportation.

Highways and the Automobile

The major component of the transportation system is under provincial not federal jurisdiction. This began in the 19th century when highways were perceived as local in nature and so came under the provinces. Roads under federal jurisdiction are still few, for example, roads in national parks and roads which are a part of federal port or airport complexes. There are only two examples of national highway programs: the Trans-Canada Highway and the roads-to-resources program of the 1960s. The effect of segmented jurisdiction among the modes of transportation is a particular issue that the Commission must address.

The early development of roads was the responsibility of the local community through the use of statutory labour. Local roads were important for property access. Early trunk roads, needed chiefly for defence and the administrative of justice, were considered a public charge. Some turnpike roads were allowed, as in the United Kingdom and the United States, but toll collection delays were of concern in the provincial road systems. Tolls were not practical for roads used dominantly for access to property.

Consequently, the pattern grew up of financing roads from general revenue, with only a share of the revenue coming from users. For many years, about two thirds was collected from users in most provinces, in the form of licence fees and fuel taxes. Fuel taxes, which provide the bulk of revenue, were

seen as user-fees varying with use, both by mileage and vehicle weight. However, the increasing fuel efficiency of large trucks with diesel power and improved engine performance have offset higher taxes for diesel fuel. In several American states, a weight-distance tax has been introduced to increase the revenue collected from large trucks.

Governments have found that fuel taxes are an excellent way to raise revenue. In addition to provincial taxes, which now approximate provincial expenditures on roads, the federal government levied an excise tax on fuel, and some local communities levy a transit subsidy tax on fuel. Total taxes paid by users in forms traditionally regarded as user-fees, now exceed highway expenditures.

Thus, taxes on fuel may have various purposes: to increase the cost of highway transportation to reflect some road costs and, thereby, to place road use, in aggregate, at a more efficient level; to act as a "pollution tax" on fossil fuel use (in transportation); to encourage a shift from automobile travel by raising its costs and, possibly, using the resulting revenue to subsidize alternate transport services; to serve as a source of general government revenue.

The latter objective appears important to the federal government. It may cause the Commission to consider the economic, social and political ramifications of policies which identify transportation, like tobacco and alcohol, as a source of high contribution to general revenue. It is interesting that the government may now see transportation as a special source of general revenue, even though, during the development of the country, special attention had been given to reducing the costs of overcoming distance. The taxes affect transportation as both an industrial input and a consumption service.

Unfortunately, the simple comparison of highway tax revenues and expenditures does not adequately compare revenues raised from users with the value of the road resources they use. Annual tax flows measure revenue; that portion (of provincial charges, especially) above normal taxes might be regarded as a user-charge. However, highway expenditures on maintenance, operations and capital are not accurate reflections of true, aggregate highway costs. They ignore the cost of capital, and do not reveal whether capital is being consumed by inadequate maintenance and replacement or being

built up by advanced construction. The 1989 study by the National Highways Policy Steering Committee suggests a substantial under-investment in highways.

The current method of financing highways is associated with a number of issues:

- What is the relationship between vehicle types and road costs?
- Would the supply of highways be more efficient if user fees were placed in a dedicated highway fund?
- Do fuel taxes result in user-fees less than or more than relevant costs in aggregate?
- How do costs and revenues compare by road type and by type of user?

As the Canadian transportation system matures, the major issue that has emerged is the efficient use of roads. The most frequently expressed concerns revolve on questions associated with the allocation of costs among users. What wear and tear are associated with which vehicle types, and do vehicle types pay for their share of road costs?

However, issues of paying for the existing system are only part of the picture. How to avoid excessive congestion? How to provide the right highway capacity? These questions are prominent because of the need for different approaches to the cost of additional highway capacity. They arise because of the general absence of road pricing — there are few toll roads.

Fuel taxes raise revenue for governments and they raise vehicle operating costs. However, they apply system wide and do not reflect the costs of specific facilities. While the appropriateness of user-fees on low-density main roads is an issue, the paramount issue for the Commission is the supply of roads to meet the demands of high traffic.

The cost of additional traffic on high-volume roads is high for two reasons. First, the traffic imposes congestion costs on the system. Second, the relevant cost of accommodating more traffic becomes not just the wear and tear on the existing highway but the cost of providing new capacity. The absence of specific road pricing systems results in the absence of the usual

measure of the need for new capacity, that is, consumers paying the price. For highways, the volume of traffic on routes is in response to the general regime of highway finance, not the cost of specific routes. The cost of new roads is much higher than the cost of existing routes. Highway users should be prepared and expected to pay the high cost of new roads.

Ensuring that highway users bear the real cost of high-volume routes is a major concern in many countries. Electronic technology is creating new opportunities to introduce road pricing. However, even traditional toll systems are being reintroduced, for example, the Coquihalla Highway in British Columbia. The economics of road pricing require careful consideration by the Commission.

Not only does the provision of additional highway capacity need much capital, it will also have important implications for the demand for public transport services. Because of the large volume of car travel, a small percentage change can have major implications for the size of public transport.

The Bus Industry

The bus industry is affected by a number of issues similar to those of automobile users. These include the level of highway taxes; the sharing of highway costs among users; and efficient approaches to traffic congestion. Intercity services are affected by urban traffic as well as trunk road conditions. However, the bus industry stands to benefit from a finding that automobile (and truck) traffic is not making sufficient contributions to counteract the cost of highway congestion or to investment costs.

The development of bus services has also been influenced greatly by provincial regulation, which remained unchanged even with the passage of the NTA, 1987. Provincial governments have licensed major carriers who then were expected to cross subsidize services to small communities with the profit from major routes. It is not evident that the program has been successful. The regulated regime may have inhibited technology and service improvements, while returns in the industry were quite high. Today, increasing competition is making cross subsidy impractical.

The position of the bus industry could be significantly affected by an integrated passenger policy dealing consistently with all modes of transportation.

The regulation of the bus industry seems anomalous. The favoured position of VIA for subsidized passenger services needs review.

RAIL SERVICES

Rail passenger services in Canada have deteriorated over the last 40 years in the face of increased competition from automobile and air transportation. In spite of occasional, well-intentioned initiatives by railway managers and governments, the contribution of rail services has declined.

Symptomatic of the problems of passenger service has been the lack of clear statements about the purposes to be served by rail services. In 1967, an amendment to the *Railway Act* provided that railways would receive compensation for 80 percent of losses approved by the Canadian Transport Commission. The incentive left with the railways to diminish the losses did not have the effects the government had hoped. CN introduced marketing initiatives which stimulated some traffic, but mounting losses and dissatisfaction led to the formation of VIA to take over the intercity passenger services of CN and CP. The passenger services provided into northern Ontario by the Ontario Northland Transportation Commission and Algoma Central continue to be subsidized under the *Railway Act*.

VIA has not been given a clear mandate, nor are the reasons clear for operating specific services. Faced with operating services at a heavy loss, VIA has attempted to increase occupancy by using discount fares which reflect the incremental cost of (subsidized) seats otherwise remaining empty. The resulting low fares led to complaints from the Montreal-based bus line, Voyageur Colonial Limited, and hearings before the National Transportation Agency. The second set of hearings was terminated following appointment of the Royal Commission. Figures such as the subsidy per passenger on routes are available but they have not been used in conjunction with detailed passenger profiles nor compared systematically with the cost of providing services by other means. While the subsidies for VIA are explicit, the reasons, effectiveness and incidence of them are obscure and muddled.

AIR SERVICES

Three types of subsidies are found in air services: incomplete cost recovery, payments for unremunerative services and cross-subsidized airline services. The main subsidies have been created by partial cost recovery.

The contributing factors to this are a scarcity of cost-recovery initiatives — for example, in connection with air navigation services; low traffic volumes at many airports; a national, uniform and inefficient system of landing fees; and limited innovation in airport development and in pricing airport services. Those matters are being addressed currently in the context of the cost-recovery initiative and airport reorganization.

It should be noted that an important issue in the proposed scheme of user-charges is the extent to which charges are designed to achieve cost recovery rather than efficiency. Other issues are whether funds should go to general revenue or an aviation fund, what role new airport authorities may play in the national system and whether air services are to be funded as a system or whether individual airports are to be treated separately.

Although aviation comes under federal jurisdiction, the provinces provide subsidies of two types. Some provinces, such as British Columbia, fund remote airports; others, such as Ontario and Quebec, support services to remote communities. In Ontario, the services are provided on an operating contract basis for the Ontario Northland Transportation Commission. They are not tendered services as would be required under the NTA, 1987.

Before airline deregulation, it was expected that national carriers, especially Air Canada, would provide some unremunerative services to small communities. Meeting this public service obligation was a responsibility that Air Canada accepted. When the forces of competition made cross subsidy impractical, the public responsibility was a concept that some managers (as well as some politicians) were slow to give up. However, innovation under deregulation has resulted in more frequent services to such small communities but with smaller aircraft matched to traffic volumes and operated by regional rather than national carriers, or by regional subsidiaries of the main carriers.

MARINE TRANSPORTATION

Large subsidies are provided for ferry services in Atlantic Canada as a result of constitutional obligations. In British Columbia, ferries are primarily a provincial responsibility, although a modest federal subsidy is provided for remote services.

The level of service and financial performance of ferry services are affected significantly by the seasonality of vacation travel. Explicit policy statements about the level of services and expected cost recovery from tourism and other major traffic segments apparently do not exist.

Where ferry services are short, as in river crossings, they may be treated as parts of provincial highway systems. These services may be provided without charge.

Although marine services can be costly for taxpayers, they do not appear to be controversial at a national level. Two conditions may account for this. First, the main need is constitutional in nature. Second, they have little effect on other modes.

IV. ASSESSMENT

Subsidy programs have evolved in the individual modes of transportation in keeping with traditional practice. No coherent approach to the roles of the modes and of the subsidies affecting them is emerging. Notwithstanding the overall policy statement in section 3 of the NTA, 1987 applicable to passenger transportation, the reality is that modal behaviour has simply evolved in a fragmented way.

MODAL SERVICES AND SUBSIDY TYPES

Subsidies are provided in transport under each of the categories recognized in Section II of this paper. However, there is significant variation across the modes. Exhibit 4 is a summary of subsidy types by mode.

Exhibit 4
SUMMARY OF SUBSIDY TYPES BY MODE

	Payment	Provision	Protection	Absorption
Auto	—	Yes	—	Yes
Bus	Minor	—	Minor	—
Rail	Yes	—	—	—
Air	Minor	Yes	(ended)	Yes
Marine	Minor	Yes	—	—

Automobile

Subsidies provided to the automobile are the least well understood. This itself is a major problem, because they are hidden in society's absorption of environmental costs and in the public provision of roads.

The technical amount and economic significance of the environmental costs absorbed by society are important issues in our time. They are slowly being addressed by environmental control measures. Such direct measures are appropriate. While the magnitude of the environmental effects of the automobile are uncertain, their existence is not doubted. The general types of response warranted are known. The environmental literature deals extensively with the role of regulations and taxes to mitigate pollution effects efficiently. Pollution is best dealt with directly, not by subsidies to alternative modes.

The most controversial highway subsidy issue is whether travellers collectively pay for highways. An equally thorny question is which parts of the highway network are subsidized. Presumably low-density roads are subsidized; this is not at debate. It can be justified on two grounds. First, these highways provide basic mobility and access, partly funded out of property taxes. Second, the subsidy can be justified economically when the cost of automobiles on roads with limited traffic is low.

The main highway issue is whether users are subsidized by the provision of high-capacity roads. This may be viewed in another way: Are users actually willing to pay for upgraded roads? (Highway users are always anxious to see road improvements which save them travel time and operating costs; they get more and pay less.) The most effective policy mechanism to deal with the issue is a move in the direction of road pricing.

Road pricing provides an opportunity to move away from the supply-driven policies of the past to those in which demand plays a prominent role. "Pay for growth" is a politically saleable principle. Electronic technology is facilitating more complex road pricing schemes than would have been possible otherwise. However, simple approaches can also work — witness toll roads and the Singapore "road pricing" system.

Effective road pricing would have the desired effect of enabling transportation costs to be seen more clearly by travellers. The commercial prospects

of public transport services would thus be enhanced. (The level of road charges in Japan and Europe certainly contributes to the viability of rail services.)

Bus Services

Bus services are provided with some intercity subsidies for unremunerative services, for example, in Newfoundland. However, the main issue in the bus industry is the hidden effects of regulation, retained, in part, to maintain cross-subsidized services. The record in other countries and in Canada suggests that cross subsidies are not sustainable, do not provide services best matched to small community conditions, and lead to less innovation and efficiency in protected services. Cross subsidization has none of the attributes of a good subsidy program, but tendered services would be possible and appropriate. New mechanisms of grants to those needing travel assistance could also be used.

Railway Services

While railway services have the merit of being subsidized mainly through explicit payments, the program is bogged down with historical legacies.

The first is our love of trains. It is necessary to recognize that the purpose of subsidies is not to subsidize trains but to subsidize particular passengers. Can better methods be found to do this? To answer this, it is necessary to have more precisely articulated statements of the objectives of subsidies and to have better information on train users. Data on subsidies per passenger are known, but who are these passengers? How many people, as compared with trips, are subsidized? How many tourists? If passengers were given the dollar subsidy implicit in rail use, what mode of transport would they use?

The second legacy is the reliance on a single main provider, VIA Rail, which is handicapped by the lack of a clear mandate. The innovative approach to tourist service between British Columbia and Alberta is evidence of the innovation possible under an alternate policy regime. If subsidy objectives are defined more precisely by route, could subsidy levels and delivery mechanisms be developed more selectively and precisely? Should tourists be subsidized to the same extent as those travelling on "essential" services? Are similar services required across markets? Should services be provided by a single supplier? To what extent can competitive tendering and service contracts be used?

Rail services have also been impeded by the hidden subsidies seemingly enjoyed by highway users in high-density corridors and by air travellers.

Air Services

Air services have had subsidies in most categories. The regulation that enabled cross subsidies for unremunerative services has ended. The environmental effects of air services are most evident in noise pollution around airports. In Canada, affected property users have, generally, not been compensated as they have been in some countries. In Japan, for example, compensation has been provided to property owners for sound-proofing buildings. In Edmonton, property owners adjacent to the city airport have been given a reduction in property taxes. However, compensation does not remove the subsidy to the air services unless the revenue is raised from them.

The noise frequency at airports is increasing because airport traffic is growing. However, the introduction of new third-generation, quieter aircraft, is greatly reducing the area of the noise imprint of flight paths and reducing noise levels within the affected areas.

Since noise pollution caused by air services continues to be dealt with by regulations on aircraft operations, some inefficiencies may result from excessive or too low standards. However, change to a system of industry payment and compensation for those affected is a change that goes beyond transportation to compensation in society generally.

The subsidization of air services through less than full cost recovery for airport and navigation infrastructure and services is being addressed in the current cost-recovery initiative. This program raises questions about the extent to which efficiency and cost recovery can be achieved; economic efficiency may require deviation from annual accounting cost recovery. The complexity of efficient user-charges has led to special studies of the subject. It has raised questions, also, about the extent to which airports should be treated as parts of a system to be financed collectively.

No air services are provided under subsidy under the NTA, 1987. However, some services to remote communities are subsidized by provinces. In Ontario, for example, the Ontario Northland Transportation Commission

has provided services through norOntair since 1973. norOntair continues to serve a mix of profitable and unprofitable northern routes, and receives an *ex ante* negotiated subsidy. Several provinces provide emergency air health-care services under contract with private operators.

Marine Services

Marine services are subsidized primarily through Crown corporations which operate at a loss. The largest losses are incurred on federal services in Atlantic Canada. The services vary by community type and traffic type, and communities vary from small and remote to large urban areas. Passengers, which include residents and business people travelling occasionally, commuters and tourists, exist in different proportions on different routes. As with rail services, the rationale and amount of subsidy incurred on behalf of the individual groups remain unclear.

Some ferry services are provided without charge as a part of a provincial highway system. While these generally are for river crossings, the apparently arbitrary distinction between services provided free and those for which there is a toll gives rise to controversy.

COMPARISON OF SUBSIDY PRACTICES AND PRINCIPLES

There is a significant contrast between the principles recommended for the management of subsidies and the attributes of subsidies found in Canadian passenger transportation. An important contribution to the effectiveness of passenger transportation in Canada would be to bring about a greater coincidence of actual with recommended practices.

The reasons for the gap between reality and recommendation are not hard to find. First, Canada started by using transport as an instrument of nation-building. Subsidies for the construction of roads and railways were to meet the basic community needs of mobility for defence and justice as well as of trade. In spite of changing needs, practices have not changed.

Second, the nature of transport services has involved government in ways that hide subsidies. These include the public provision of facilities and services, as in public roads, airports and ports, and in the regulation of for-hire services.

Third, the limited role of competitive markets has restricted the influence of commercial pressures. Disciplines of the market have been stronger in the freight mode.

Finally, politicians have an understandable but unfortunate reluctance to reveal the level and incidence of subsidies for fear of upsetting their constituents.

The desirable attributes of subsidy programs are:

- Provide subsidies as directly as possible to the intended beneficiaries;
- Provide subsidies in ways that do not distort competition;
- Avoid subsidies which distort input prices;
- Make subsidies transparent, that is, visible and with a clear purpose; and
- Monitor and publicize results.

The most important of these attributes is that subsidies be transparent, a quality that is essential for effective monitoring. Clarity of purpose must go beyond statements of provision of service to a particular community and measures of total ridership, revenues and costs. It must identify particular types of users and beneficiaries.

Detailed technical studies on the effectiveness and incidence of specific passenger service subsidies would be pathbreaking. They would make contributions comparable to the grain costing work of the MacPherson Commission, which provided an example of technical analysis and provided a basis for important policy recommendations.

The attributes of desirable programs do not point directly to one type of subsidy over another. However, the nature of subsidies as they exist in Canadian passenger transportation suggests that those which are least likely to meet the desired criteria are absorption and protection, followed by subsidies through the provision of facilities and services without full cost recovery. The payment method of subsidization is likely to be able to meet the criteria. Unfortunately, as practiced in Canada at present, it does not do so.

The absorption of costs is undesirable as it encourages the inefficient use of resources. In transport, it is mainly associated with environmental effects which are being dealt with directly, albeit late and slowly.

Subsidies by protection, such as in bus services, are hidden, often ineffective for those subsidized and result in less efficient services overall. Subsidized services would be more effectively provided if service contracts or competitive tendering arrangements were used. The extent of competitive tendering may be influenced by the market structure and policies respecting Crown or private operations. Increasingly, the latter is preferred and competition more likely. The provision of direct “payment” to travellers may also be considered.

The provision of infrastructure by governments creates situations in which subsidies may be hidden and purposes unclear. Less than full cost recovery for low volume airports or roads may be efficient and consistent with other policy objectives.

Subsidies may exist also in the provision of intensively used facilities. Congestion costs and the high cost of new facilities must be considered. The similarity of the economically correct policy of having users pay marginal costs with the politically acceptable policy of users “paying for growth” may provide an important basis for policy recommendations. This principle may be particularly helpful in advancing the case for more direct charges than fuel taxes for roads with high traffic volumes. Various schemes are making road pricing a realistic option for the future.

User-fees designed to raise the level of cost recovery will likely raise concern for the efficiency with which facilities are currently used. That is good. For example, the increase in user-charges for air and sea modes has raised awareness about the inefficiency of supply of facilities for those modes. A similar response from road users would probably increase interest in the more efficient use of road space, for example, by high-occupancy vehicles.

The development of new policies for user-charges raises questions about the use of dedicated modal funds. Such funds would not be recommended in theory, but perhaps they are desirable in practice. It should be noted that the discipline that such funds create may not be liked by politicians!

The payment of subsidies to VIA is the most visible passenger subsidy in Canada. The program does not meet a number of criteria for a desirable subsidy program. The target travellers are not identified clearly so that

results cannot be measured effectively. The subsidy distorts competition among modes by being provided to a supplier, VIA, rather than to travellers. Competitive tendering of the transportation services does not take place.

Overall, the provision of subsidies in Canadian passenger transportation has evolved over time and under different governments without a coherent framework to ensure that the subsidies would work to achieve a safe, adequate, economic and efficient transport system making the best use of all modes of transportation. A significant shift in subsidy practice is needed to achieve the overall policy goals.

Current practices and policies reflect the historic need for governments to ensure the provision of basic infrastructure and services to build a new nation. Attitudes and institutions have not yet adapted to the requirements of a mature nation in which resources must be allocated among alternate programs, including coping with congestion and expansion. The implication is that less attention should be given to engineering-driven supply initiatives and more attention given to demand-side elements. This includes a stronger role for market mechanisms, such as cost recovery, including road pricing, and a greater attention to particular travellers for whom assistance may be warranted. The desirable attributes of subsidy programs outlined previously are consistent with these recommendations.

Passenger transportation policies in the future need to provide a long-term vision which can guide the development of promotional policy based on the requirements of a mature economy. This shift will be as significant as the change achieved in regulatory policy since the MacPherson Commission first recognized the need to respond to the changing role of competition in transport markets.

ENDNOTES

1. Organisation for Economic Co-operation and Development, *Transparency for Positive Adjustment: Identifying and Evaluating Government Intervention*, Paris, 1983, pp. 32–33.
2. Ibid., p. 249.
3. Rachel E. Kranton, *Pricing, Cost Recovery, and Production Efficiency in Transport: A Critique*, Infrastructure and Urban Development Department, The World Bank, Working Paper 445, Washington, D.C., 1990, p. 45.
4. Mark W. Frankena, "Capital-Biased Subsidies, Bureaucratic Monitoring, and Bus Scrapping," *Journal of Urban Economics*, 21 (1987), pp. 180–93.

TRANSPORTATION INFRASTRUCTURE POLICY: PRICING, INVESTMENT AND COST RECOVERY

David Gillen* and Tae Hoon Oum**

May 1992

1. INTRODUCTION

Transportation policy issues were, for much of the last several decades, loosely integrated with national and regional economic policies in Canada. Transportation was viewed as a tool of government to promote the economic development and growth of different regions, groups and industries. This position changed somewhat with the passage of the *National Transportation Act* (NTA) of 1967. Based on recommendations of the MacPherson Royal Commission, the Act broke with tradition and established the policy that efficiency within the transportation sector was of fundamental importance in securing a firm foundation for the economic growth and development of Canada. To achieve this goal, competition among modes was selected as the mechanism. The MacPherson report also recommended that any region, group or industry deemed sufficiently deserving of subsidization was to receive direct payment rather than subsidization through the underpricing of either or both transportation services and infrastructure. This was generally adopted as part of the NTA of 1967.

* School of Business and Economics, Wilfrid Laurier University and Institute of Transportation Studies, University of California, Berkeley.

** Faculty of Commerce and Business Administration, University of British Columbia.

Since 1980, external and internal economic and political pressures have resulted in deregulation in the airline and trucking industries, the introduction of intra-modal competition and the privatization of carriers. This policy shift was formalized in the new *National Transportation Act, 1987* which took effect January 1, 1988.

Allowing market forces to determine the structure and behaviour of different modes (primarily carriers) has been the new policy direction. It has also meant a shift away from setting infrastructure prices below costs. Transport Canada's recent proposal of a cost-recovery policy for air and marine infrastructure is a small step toward this goal. A similar shift has not taken place for highways, likely because of the fractured jurisdictional responsibility for roads among the different levels of government. While air is the responsibility of a single agency (the federal government) highways are a combination of provincial, regional, county and municipal responsibilities with the proportions of roadway in each of these levels of government varying across provinces. This spread also characterizes the United States and distinguishes North America from Europe where there is more integrated transportation planning at the central level.

Over this same period, infrastructure policy has been subjected to a number of exogenous stresses including the Free Trade Agreement with the United States and the globalization of industries. Both factors have resulted in growth and shifts in demands which have generated pressure for additional and improved infrastructure as well as for new institutions to manage them. These demands come at a time of restraint with governments under severe economic pressures to be fiscally responsible. Governments are reluctant to commit themselves to significant expenditures even for maintaining the existing stock of transport infrastructure let alone adding to it.

Given the realities facing the Canadian economy, how should one establish pricing principles and investment guidelines which satisfy both the objectives of the *National Transportation Act, 1987* and the needs of Canada and the Canadian economy in the years to come? The fact of the matter is that transportation infrastructure, an important component of the nation's capital stock and an important factor in its economy and welfare, is a victim of both fiscal restraint and the failure to treat it as a scarce economic resource. In the past, transportation policy at all levels of government focussed on capacity expansion (supply) rather than the management of demand for

infrastructure. Sensible policy reform must consider the costs imposed by users on the system and on others when using infrastructure, including congestion and other externality costs. Infrastructure shortages resulting from both underpricing and inadequate investment limit the realization of the efficiency gains promised by deregulation and privatization in the transportation sector.

Pricing and investment planning of Canada's transportation infrastructure cannot ignore the forces and pressures developing in the United States and other international markets. Links must be established between investment in infrastructure and the pricing of services delivered by that infrastructure. Indeed, socially optimal modal pricing requires the inclusion of modal air and noise pollution and congestion externalities. Congestion externalities are emphasized in this study but the models and concepts are applicable to air, noise and other externalities. Economic welfare will be lower if these factors are not considered in modal prices because demands for massive public investment in infrastructure will continue unabated, infrastructure will deteriorate prematurely, and the distribution of traffic across modes will not reflect the real modal costs. The solution is not necessarily more investment but rather smarter investment. Smarter investment must start with efficient pricing.

The main purpose of this study is to discuss the principles and methods by which Canada can ensure the optimal use and efficient provision of transportation infrastructure services in all modes of transportation. Emphasis is also placed on the cost-recovery issue. There are a number of reasons for this, including the fiscal constraints faced by government, the motivation to reduce the efficiency costs of financing infrastructure deficits through taxation and the general movement to decentralization.

Cost recovery is treated in two alternative ways. First, cost-recovery conditions associated with optimal pricing and investment in infrastructure are studied and compared to the actual cost-recovery situation. Second, the methods of achieving an exogenously given cost-recovery target which minimizes efficiency loss is discussed. Although the principles and methods used can be applied to all modes of transportation, the emphasis of the discussion (with empirical examples) is on roads and airport infrastructure. The study also emphasizes that allocative and productive efficiency gains are not accomplished through simply financing a cost recovery but rather through achieving cost recovery with efficient pricing.

Section 2 presents a survey and assessment of the principles and alternative methods of infrastructure pricing. It includes a discussion of the potential difficulties for implementing some of the ideal principles. Also reported are the empirical results of some studies which apply the optimal pricing principles. Section 3 examines the literature on cost structures for transportation carriers and infrastructure providers. The focus is given to the empirical results on economies of scale, traffic density and scope. An analysis of the optimal user-charges for airports and roads is presented, and some suggestions for achieving the optimal cost-recovery targets are made in Section 4. This section also reviews current user-charges and cost-recovery status by major user-groups in both air and road modes. Alternative sources of financing road infrastructure currently in use are examined and discussed in Section 5. A summary of the findings is given in Section 6.

2. INFRASTRUCTURE PRICING: THEORY, ISSUES AND APPLICATIONS

This section presents a survey and assessment of the principles and alternative methods of infrastructure pricing. Although it addresses infrastructure pricing principles, the characteristics and consequences of each pricing principle will also be discussed in a general framework for all goods and services. Potential difficulties in implementing some of the ideal principles are identified and discussed. The empirical results of some studies which apply the optimal pricing principles are also presented.

2.1 THE ROLE OF PRICES AND WELFARE

Transportation infrastructure is like any piece of capital — it represents a stock which yields a flow of services. The services from infrastructure, such as roads and airports, reflect economic scarcity because their construction and continued servicing require the use of scarce resources. Concern about the efficient use of resources in the economy necessitates finding some mechanism and criteria to determine both the distribution of resources between transportation infrastructure and other parts of the economy and resource distribution among modes within the passenger transportation sector. How many miles of roads should be built? How many airports? How big should they be? These, among others, are questions which pricing principles seek to answer.

There are essentially two perspectives as to how to answer these questions. One is the market approach. It uses the price system to allocate scarce resources to their most valued uses on the basis of willingness to pay. If the market mechanism is to be used as a method of maximizing social welfare, all private and public goods and externalities (such as air or noise pollution and congestion) must be "valued." This value or price should reflect the resources used and be a measure of the cost to produce the good or service. This cost is to be the marginal cost, and no buyer or seller of a good or service should have sufficient power to drive a wedge between the price charged and the marginal cost of production. These conditions assure maximization of allocative and productive efficiencies or, equivalently, of social welfare. Allocative efficiency is the measure of performance of scarce resources allocated to end uses, goods and services, that best accords with the pattern of consumer demand. Allocative efficiency is at an optimum when the price of each product equals the lowest resource cost of supplying the marginal unit of the product. Technical or productive efficiency refers to the minimum cost of producing a given output.

The other approach to resource allocation is to use direct planning methods characteristic of some centralized economies and which include the use of various non-market mechanisms. These non-market mechanisms can take a variety of forms including administrative rationing, random allocation and queuing. Some authority must decide how to allocate available goods or services to competing uses and choose the goods and services to produce. However, since there is no efficient method of inducing users to reveal their preferences, the authority has no way of accurately knowing who values a given good or service more. Consequently, the authority may end up providing a good or service to those who do not value it the most. This can lead to significant allocative inefficiency. In most cases of direct market intervention, the objective is rarely to improve efficiency. Efficiency costs, however, are still important as measures of the costs of abandoning a market approach. Furthermore, because there are no signals for the capital market, there will not necessarily be an optimal investment in capacity, except by chance.

In a market economy, prices perform two functions. In the short run, they act as a signal to ensure that scarce goods and services (airport capacity, for example) are allocated to those who value them the most. This ensures that the social benefit, from the utilization of the fixed capacity, for example, is

maximized. In the long run, prices provide a signal to the capital market to move capital into those activities which yield the highest return, and thus guarantee an optimal investment in capacity.

Given the different outcomes associated with the alternative policy proposals, efficient pricing or non-market mechanisms for resource allocation, one must use some criteria to assess their relative merits. Economists have generally not considered both economic welfare *and* income distribution outcomes when evaluating alternative policy proposals or pricing methods. The level of economic welfare is defined as the sum of consumer surplus and producer surplus. Consumer surplus is the additional value a consumer derives from consuming a good or service over and above the price paid and is generally measured by the difference between the value revealed by the demand curve and the price paid. Producer surplus is a producer's net revenue over and above the cost of production. A dollar saved by a low-income person is regarded as having exactly the same value to society as a dollar saved by a high-income person. This means evaluation of alternative policy proposals or pricing methods by the welfare criterion alone ignores the income distributional consequences. In other words, an economic reorganization or change is considered beneficial if those who benefit collectively gain more than the total losses incurred by those who lose. Use of this welfare criterion alone is based on the compensation principle which implies that those who gain can compensate those who lose without incurring any redistribution costs. Although the change in the level of welfare (economic efficiency) is generally regarded as more relevant for evaluating policy, the effect on income distribution cannot be ignored because, for most cases, income redistribution is not without cost. The various pricing principles and non-market allocation procedures reviewed are judged primarily in terms of their consequences on economic efficiency (welfare). Attempts are made, however, to make a preliminary assessment of the consequences of some pricing principles for income distribution (fairness).

2.2 ALTERNATIVE PRICING METHODS

Pricing is a method of allocating resources. There is no such thing as the "right" price irrespective of issues and objectives. Rather there are "optimal prices" or pricing strategies with particular objectives to be achieved. Prices can be established to maximize profit, welfare or revenues. They can be used to achieve a particular share of the market or a desired distribution

of demand across products (for example, mode-split in transportation). From a society's viewpoint, however, one of the most important goals for pricing goods and services is to maximize economic welfare by optimally allocating scarce resources and goods or services across competing needs in the short run and to ensure optimal investment in capacity in the long run.

Much of the discussion and debate surrounding the pricing of runways and roadways seems to confuse efficient pricing and cost recovery or financing. Pricing transportation and financing transportation are very different concepts. Financing requires only knowledge of costs, and user-charges are set to achieve full or a desired level of cost recovery. Pricing, on the other hand, requires knowledge of both demand (which provides a measure of economic value) and costs since pricing tries to optimize the use of the resource and to balance the revealed value of transportation with the resource cost of providing it. In analyzing various pricing strategies it is, therefore, essential to have a clear understanding of various cost concepts since each pricing principle is related to a specific concept of cost.

Before proceeding with the detailed discussion of the various pricing methods, it is important to clarify the use of subsidies from general revenues to meet revenue shortfalls under some pricing approaches. There are two points. First, it has been argued that any deficits resulting from marginal cost pricing be financed from general tax revenues. Second, it is also argued that full cost recovery is not desirable if it results in prices that do not reflect marginal costs. First-best pricing has prices equal to marginal cost and second-best pricing has prices which deviate from marginal cost in a way which minimizes efficiency losses. Both arguments have an implicit assumption that the efficiency loss is less from raising a dollar for general tax revenues than by raising revenues in a specific market. This is not necessarily true. Jorgenson (1992) noted that the cost of public funds is indeed high. He reported on research by Jorgenson and Yun (1990) which shows that the marginal cost of a tax dollar is \$1.46. In other words, for every dollar of public spending, the cost is \$1.00 in tax revenue and 46¢ of loss in efficiency of the private sector. Ballard, Shoven and Whalley (1985) produce a marginal cost of \$1.33. These numbers seem to justify a lower level of subsidy to some transportation modes than is implied in the current transportation literature. In essence this means that, with a subsidy from general revenues, second-best pricing may result in a lower efficiency loss than does first-best pricing.

The following subsections describe various types of pricing which may be relevant for transportation infrastructure pricing. These are: average cost pricing, marginal cost pricing, social marginal cost pricing, Ramsey quasi-optimal pricing, peak/off-peak pricing and multi-part pricing. They are followed by a subsection on allocation by non-market mechanisms, such as slot allocations at airports.

2.2.1 Average Cost Pricing

The average cost price is obtained by taking total costs and dividing them by the relevant measure of output. For example, if total annual airfield costs are \$1 million, and if 10,000 landings per year are expected, then the landing fee would be set at \$100. If price is set equal to average cost, total costs will be recovered. Average cost pricing is sometimes termed “full cost pricing.”

If an industry’s production technology is not characterized by constant returns to scale over the relevant range of output, average cost pricing leads to economic inefficiency by making the value (to society) of producing another unit deviate from the cost of producing it (that is, marginal cost). It is easy to demonstrate that if average cost is falling (increasing returns to scale), setting price at average cost lowers economic well-being due to the underproduction of services. Producing an additional unit of output at a cost equal to marginal cost and selling it for a price that exceeds marginal cost (but is less than average cost) could provide an increment to profit and potentially make society better off. Average cost pricing prevents such a desirable action from occurring.

Unless constant returns to scale prevail in the relevant range of market demand, average cost pricing leads to an incorrect level of output, both in the sense of social well-being and in the sense of profit maximization. Thus, except in the special case of constant returns to scale, average cost pricing is not a desirable basis for establishing a pricing strategy.

2.2.2 Marginal Cost Pricing

Marginal cost pricing maximizes the economic benefit to society by ensuring that a socially optimal volume is traded in the marketplace, and the optimal quantity is allocated to those who value it the most. Since marginal cost pricing maximizes social benefit without any constraint, it is often referred

to as "first-best" pricing. For transportation services, this means that the service output is extended to the point where the marginal cost of serving an additional unit equals the price the user is willing to pay for the service. This marginal cost can be quite different from the average cost of production. Suppose for example, that the total cost of servicing 100 aircraft movements is \$1,000 and servicing 101 movements is \$1,005. This means the average cost of servicing 100 movements is \$10, but the marginal cost of providing an extra movement (when the current service level is 100) is only \$5. Average cost pricing would set the price at \$10 per movement, while marginal cost pricing would set the price at \$5 (for all movements). This pricing principle resolves the problems of inefficient production and consumption associated with the average cost pricing principle. Under conditions of constant returns to scale, unit costs are not rising or falling in the relevant output range. In this case, marginal cost equals average cost. Thus, marginal cost pricing and average cost pricing would be the same.

When unit costs are either rising or falling, average and marginal cost pricing strategies produce different results. Preference is generally given to marginal cost pricing based on the assumption that the efficiency loss resulting from any deviation from marginal cost pricing is greater than the efficiency loss of raising revenue through taxation. While average cost pricing would always lead to a financial break even, marginal cost pricing would require a subsidy for a firm to break even in the case of a declining unit cost industry. This is because price (that is, marginal cost) will be less than average cost. Marginal cost pricing will result in break even in the case of constant cost and a profit in the case of rising unit cost.

Another problem with marginal cost pricing is in the difficulty of measuring appropriate marginal costs. In the case of average cost pricing, one simply sums all relevant costs, operating and capital, variable and fixed, direct and indirect, and divides the sum by the anticipated output level to obtain price. For marginal cost pricing, however, it is not easy to identify the costs which vary with output even for the case of a single product. For instance, capacity costs may be fixed in a very short run but may vary over a longer run. Therefore, the measure and variability of marginal cost will depend on the time frame one chooses. This study argues, and this is the consensus in the literature, that the short run is the appropriate time frame for establishing efficient prices.

2.2.3 Social Marginal Cost

When the production or the use of goods or services results in negative (positive) externalities, such as the imposition on others of noise or air pollution or congestion delays, social marginal cost deviates from private marginal cost (cost incurred by the producer or user). In this case, marginal cost pricing based on the private marginal costs will generate too much (too little) output compared to the socially optimal level of output. When this occurs, social marginal cost pricing maximizes welfare (economic efficiency) as it internalizes the externality costs in a user's decision making. Social marginal cost pricing extends the marginal cost pricing principle to the situation where externality costs exist. It recognizes the costs a user imposes on others (externality costs) in addition to the privately borne costs of the user. For instance, an extra vehicle that uses a congested highway imposes costs on the cars and drivers following it by imposing additional delay on them. It also imposes air and noise pollution. In the case of airports, a user of the airstrip may impose a congestion externality, as well as noise and air pollution costs. Social marginal cost pricing internalizes these externality costs by charging users the full social cost. By internalizing the externality costs, the user is induced to make decisions consistent with social benefit maximization.

The application of this pricing principle has been shown to result in the optimal use of a facility (given fixed available capacity) and in the optimal level of facility investment (see, for example, Morrison, 1983). From as early as 1920, this principle was presented in the context of transportation, specifically road congestion (Pigou, 1912; Knight, 1924). In 1961, Walters formalized the peak-load pricing work of Boiteux (1960) and Steiner (1967) using cost functions. About the same time, Strotz developed the same ideas but with the use of utility functions. The basic model has been elaborated upon and extended by numerous researchers including Mohring and Harwitz (1962), Vickery (1965, 1968), Mohring (1970, 1976), Keeler and Small (1977), DeVany and Saving (1980) and Jordan (1983a, 1983b). Morrison (1986) showed that the conventional homogeneous users model needs only slight modification to handle heterogeneous users.

2.2.4 Ramsey Quasi-Optimal Pricing

As discussed earlier, there are circumstances where marginal cost pricing would not be sustainable in the long run. For example, if the production function is characterized by increasing returns to scale, marginal cost is less

than average cost and, therefore, total revenue would be insufficient to cover total cost. In this case, there are generally three options to choose from:

- use marginal cost pricing in conjunction with a government subsidy;
- use some form of second-best pricing, such as Ramsey pricing; or
- use multi-part pricing.

The features of Ramsey pricing are described below. Multi-part pricing is considered later.

Economists have demonstrated that a deviation from (social) marginal cost pricing may reduce social benefits and misallocate resources by over or underproducing the service. The essence of Ramsey pricing is to minimize the loss of economic efficiency caused by the deviation of prices from their respective marginal costs while allowing a financial break-even position to be achieved. Ramsey pricing makes use of an inverse-elasticity rule to mark prices up over marginal cost while ensuring that the quantity of service supplied deviates by the least amount from the optimal quantity under marginal cost pricing. In doing so, Ramsey pricing makes use of the willingness of the segments to pay. In other words, Ramsey pricing maximizes social welfare subject to the constraint that the firm achieve a break-even financial position. Thus, it is usually referred to in the literature as a “second-best” strategy.

The Ramsey pricing principle states that when a revenue constraint exists, the ratio of the markups (the excess of the selling price over marginal cost) must be proportional to the inverse of the price sensitivities for the product in question. In other words, the different groups of users may pay different prices depending on their own price sensitivity, even for an identical product or service. If applied to airfields, this might imply that landing fees would be dependent on the length of the flight, aircraft size or type of use because each of these (demand) characteristics can be expected to lead to different price elasticities of demand. A simple approach to covering the fixed costs is to charge all types of users an equal price which exceeds the price based on marginal cost pricing. Although simple, this approach is inefficient and leads to a larger loss of social benefit than the allocation based on Ramsey pricing. According to the Ramsey pricing principle, these fixed costs are recovered by allocating proportionally more of the fixed costs to those who have a lower price elasticity of demand than those with a higher elasticity.

Ramsey pricing relies on the existence of different market segments with a different willingness to pay for the same good or service.¹ Ramsey prices must cover variable costs. It minimizes the total loss of social welfare by allocating the excess of total costs over variable costs to various market segments. For Ramsey pricing to achieve social optimality, it is essential for the producer to have some market power; otherwise, the set of prices established under a Ramsey rule cannot be sustained.²

Ramsey pricing can be generalized as a method of finding the optimum set of prices under any revenue constraint. This constraint can take the form of being confined to a given level of subsidy, a requirement to break even, or even to achieve a surplus of revenues over costs of a given magnitude.³ Whenever the constraint becomes non-binding, the set of optimal prices derived from Ramsey pricing will become identical to those emerging from marginal cost pricing.

2.2.5 Peak/Off-Peak Pricing

Peak-load pricing is a widely used method in pricing public utility services. Peak-load pricing means that peak-period users are charged a higher price than off-peak users. First, peak-load users impose higher costs on the service provider than do off-peak users because they generate the need for capacity expansion, and thus should be prepared to pay all of the capacity costs. Second, peak-period users may also impose congestion costs and thus should be charged their social marginal costs. Third, since peak-load users generally have less price-elastic demands than off-peak users, charging higher fees to peak-period users (who value the service the most) is consistent with the spirit of Ramsey pricing. In other words, Ramsey pricing serves as a rationale for peak-load pricing.

Whatever the economic justification chosen, the end result is essentially the same: charging higher fees to peak than to off-peak users. Charging higher prices to peak users enhances economic efficiency by inducing them to make rational choices as well as helping solve financing problems for capacity expansion. Finally, peak pricing is not, in principle, unfair or inequitable. It assigns costs to those who are responsible for them. It makes no economic sense to restrict the use of a facility in off-peak hours because all that results is underutilization of an existing facility. With socially efficient pricing, peak users are no worse off in terms of what they pay, provided there is a low off-peak price, and off-peak users pay at least their variable costs.

2.2.6 Two-Part Pricing⁴

Another approach which has been developed to price infrastructure and to recover fixed (including overhead) costs is to have a multi-part price. One part, an entry fee for access to a facility (or infrastructure) is fixed and confers the right to use the facility. A second part, a usage fee is a price per unit of use of the facility. A price per kilometre would be an example.⁵ The intuitive appeal of the two-part price is the ability to distinguish between the value of the potential demand for access to use and the demand for the actual use of a product or network. The prices of access and usage must be correlated with the proportions of fixed and variable cost while at the same time satisfying a revenue constraint. A person can be charged an access fee even if there is no use of the system. For example, people may benefit from having a roadway because others can visit them. They may at some future date decide to use it, and it may reduce transactions cost to have continual access to a facility rather than to contract for access each time they choose to use it. The ability to match prices with demand, representing different valuations of access and of usage, results in a higher level of economic welfare than if only a single price is charged. Train (1991) summarizes a broad literature which demonstrates that a tariff structure with $N+1$ tariffs *will always* Pareto dominate a structure with only N tariffs when price exceeds marginal cost and marginal cost differs from average cost.

With a two-part tariff the level at which each part is set (and the ratio of the revenues from each of the two parts) depend on a number of factors. If the demand for access is completely inelastic, the ideal approach is to set the access fee equal to the fixed cost and the usage fee equal to the marginal social cost of use. In effect, the access fee acts like a lump sum tax. When access demand is price sensitive and therefore not independent of the level of the access fee, the access fee cannot be treated simply as a mechanism to cover fixed costs. Raising the access charge involves a loss in consumer surplus as well as revenue from usage because some consumers will choose to forego access to the service (phone line or vehicle licence, for example) in the face of the higher access price. First, as the access charge rises more users are foreclosed from the market. This reduces consumer surplus and revenue from the access charge, if access demand is at all elastic. Second, with fewer users the demand for usage declines; the demand curve shifts to the left. This means a lower level of consumer surplus and revenue from usage. This fact must be incorporated into the determination of optimal prices.

When the number of users of the network, facility or system is affected by the level of the access fee, the institution or authority which has been charged with the responsibility of setting the prices must balance the relative welfare gains and losses as the relative access and usage fees are raised or lowered. If the access fee is lowered to encourage more users, the usage fee must be increased to compensate for the loss in revenue. The amount by which it will have to be increased with the objective of maximizing economic welfare while covering all costs (breaking even) increases with the absolute access price elasticity of the number of users and decreases with the absolute usage-price elasticity of the amount of usage (Ng and Weisser, 1974).

The absolute and relative values of the two elasticities are important in determining the levels of the access fee and usage fee. The access fee does limit the number of users, and it covers some or all of the fixed costs. The usage charge allocates facilities while covering the variable and part of the fixed costs. If the ratio of fixed to total costs is high, it is desirable to have as many members as possible implying a lower access fee and a usage charge greater than usage marginal costs. However, this cannot continue unabated since as the variable charge is increased above variable costs both revenue and welfare are reduced. The ratio of revenue contributions from the variable charge to the fixed charge depends on both the relative and absolute values of the elasticities of access and usage.

When access demand is price sensitive and a financing constraint is in place, the Ramsey concept can be used to compute the second-best tariffs for access/usage services. The Ramsey rule leads to setting the usage and access fees above their marginal cost. Train (1991) derived the following Ramsey rule for determining the optimal access and usage fees that allow the firm to break even.

$$\frac{P_a - MC_a}{P_a} (\epsilon_a - \epsilon_{ua}) = \frac{P_u - MC_u}{P_u} (\epsilon_u - \epsilon_{au}) \quad (2.1)$$

In the equation a refers to access and u refers to usage. ϵ_a is the elasticity of demand for access with respect to the access fee, ϵ_{ua} is the cross elasticity of demand for usage with respect to the access fee and ϵ_{au} is the cross elasticity of demand for access with respect to the usage fee. The rule in this situation states that the percent by which the access fee is raised above the marginal cost of access, multiplied by the "net" elasticity for the access fee,

is equal to the percentage markup of the usage fee multiplied by the "net" elasticity of demand for usage. The practice of setting the access fee equal to zero and increasing usage fees above marginal cost, as is done for example at airports and electrical and gas utilities, is optimal only if the demand for usage is fixed and the marginal cost of access is zero. Usage demand is usually more price sensitive than access demand because usage is conditional on access. Some economic efficiency gains can, therefore, be expected if there is a move to some reliance on the access fee (Train, 1991).⁶ Such a change has been introduced in New Zealand in the pricing of air traffic control services.⁷

A two-part tariff can be used if users are relatively homogeneous. It is possible, however, that a facility, such as a roadway or an airport, may have a number of groups of users and that preferences may vary significantly across groups of users. It may, therefore, be desirable to have the two-part tariff and access/users fees vary for these different groups. For example, some groups may have low access and high usage fees and others high access and low usage fees. Generally, welfare is improved by offering consumers a menu of choices of two-part tariffs.

In sum, when access demand is price sensitive, the optimal access fee is lower and the optimal usage fee is higher than when access demand is fixed.⁸ Unlike the situation with fixed access demand, the first-best outcome is not attained when access demand is price sensitive and the firm is required to break even. The reason for this is clear. When access demand is price sensitive, the access fee cannot serve simply as a subsidy mechanism since it also affects access demand and, indirectly, usage.⁹

2.3 NON-MARKET MECHANISMS

In the absence of pricing mechanisms, there are non-market or administrative instruments which are used to allocate scarce goods and services and to undertake new investment to expand capacity. A variety of these have long been used for the allocation of goods, services and resources. These include administrative rationing, random allocation by lottery and queuing. Despite the variety, all non-market mechanisms share the common characteristics that they do not use prices to allocate resources, goods and services. The outcome of both the resource allocation and the level of investment

bears little resemblance to what would occur with efficient pricing. Measured in terms of economic efficiency, such allocative mechanisms are not generally as good as those obtained through the market system.

In general, non-market methods lead to inefficient outcomes due to their inability to adequately distinguish between high- and low-valued uses. The prices individuals are willing to pay directly signal such values. The inability to distinguish between high- and low-valued uses often leads the non-market methods to allocate some resources to those users who do not value them as much as others or who do not value the product or service as much as it costs. This results in a reduced level of welfare for society as a whole. Another problem with non-market mechanisms is that there is no built-in pressure which signals optimal timing and amount of capacity expansion. As a result there may be too little or too much capacity. Finally, the lack of a "market discipline" can, and generally does, lead to higher costs than would otherwise be the case.

Administrative allocation has two additional problems. First, the allocative principle tends to be arbitrary and could change with the decision-making authority and the political climate. Second, the resulting allocation of slots or rights to use a facility at a given time could create some monopoly power in secondary markets (such as airline services) by restricting new entrants especially during peak times. In the case of market allocation, such discriminatory allocation problems do not arise as the prices are set to clear the market and optimal capacity is likely to be in place. The problem of restricting potential entrants becomes particularly serious when the allocation committees are controlled by the existing users as in the case of the airport slot allocation committees at major Canadian airports.

Social welfare can be enhanced by allowing slot sales after the initial allocation. Welfare is improved because those who value particular slots the most end up using them. However, this results in windfall income gains to those who are initially allocated the high-value slots. This creates a severe inequity problem. An alternative is to let the airport auction its slots so that it gets to keep the windfall income gains.

Either type of slot market also has the advantage of signalling to the airport when additional capacity should be added. When the slot price is greater than the cost of incremental capacity, then investment will make society

better off.¹⁰ Even competitive slot auctions, however, do not always guarantee that social marginal costs are being charged as bidders ignore externality costs.

2.4 POTENTIAL DIFFICULTIES OF IMPLEMENTATION

Thus far various pricing methods which may be applied to pricing transportation services and/or infrastructure services have been presented. Some potential and practical difficulties of implementing some of the pricing methods are identified and discussed in the following subsections.

2.4.1 Pricing a Single Mode Versus Several Modes

All transportation services are capable of being supplied by more than one transport technology or mode. Each mode has different costs and quality characteristics. One of the aims of transport policy is to implement appropriate charging schemes for different modes in order to maximize the overall economic efficiency of the transportation sector. For example, if certain buses, which compete with railways, pay more than they should for the costs they are responsible for, such as the use of highways, then this may lead to an inefficient allocation of passenger traffic between bus and rail modes. An efficient allocation of traffic among competing modes requires marginal cost pricing by all modes as this induces users to make “socially optimal” choices among competing modes.

Problems arise if marginal cost pricing results in deficits for one or more modes of transportation. In such a case, each mode can be subsidized from general tax revenue or, in principle, the combined deficit of all modes can be allocated using the Ramsey pricing method. The former solution may be impractical as governments try to become fiscally conservative. On the other hand, it is difficult to implement the latter as the application of the Ramsey method intermodally requires the centralized control of all transportation infrastructure by a single agency. Currently different levels of governments have jurisdiction over different modal infrastructure.

Even if various governments were to cooperate for the implementation of Ramsey pricing in a multimodal context, it would require knowledge of each mode’s combined cost of modal services (for example, airline services) and infrastructure (for example, airport services) as well as price elasticities

of the demands. To date, there is no definitive empirical work which examines the structure of the combined cost of service and infrastructure for any mode. This makes it difficult to actually compute optimal prices which maximize economic efficiency in the multimodal context (see Oum, 1981; Winston, 1985 for imperfect attempts for such computations).

2.4.2 The Indivisibility Problem — Lumpy Capacity Expansion

Typically, economists assume that capacity is divisible when they investigate optimal pricing, the provision of capacity and cost recovery issues. For example, Mohring and Harwitz (1962) and Mohring (1970, 1976) showed that optimal congestion toll revenue exactly equalled the capacity investment cost when highway construction (and expansion) was perfectly divisible *and* was characterized by constant returns to scale. The debate regarding the divisibility of capacity expansion, however, remains unsettled, and various theoretical results obtained under the assumption of perfect divisibility require some refinements for cases where capacity expansion is characterized by lumpy investment, such as airport terminals and runways, and expressways.

There are two opposing views in the road pricing literature concerning the divisibility of capacity expansion. Some economists believe that increases in road capacity are characterized by lumpy investments (Walters, 1968; Kraus, 1981; Starkie, 1982) while other economists argue that the capacity of a road can be expanded relatively smoothly by simply improving some design features or by improving the traffic control system. Kraus (1981) showed that the cost-recovery theorem of Mohring and Harwitz required modification under the condition of indivisibility of capacity construction and expansion. The financial performance of roads under optimal pricing and investment policy depended on the type of road: that is, high-capacity roads or low-capacity roads. He found that, in the case of major highways, the cost-recovery ratio was higher when indivisibility was taken into account than for the case of perfect divisibility.

Oum and Zhang (1990) investigated the long-run relationship between congestion toll revenue and the capital cost of an airport. Their model incorporated the lumpy nature of capacity expansion and the demand fluctuations within a given day and over time. They showed that the cost-recovery ratios realized under social marginal cost pricing and optimal investment depend

on both the time pattern of traffic growth and the amount of initial capacity in place. The most significant of all empirical results, developed for Pearson International Airport, Toronto, was that the larger the existing capacity, the higher the ratio of congestion toll revenue to capacity expansion cost.

The assumption of divisibility of capacity construction and expansion affects the setting of cost-recovery targets and user-charges. The relevant empirical question is to what extent the less-than-perfect divisibility of capacity expansion makes the cost-recovery results deviate from the classic result of Mohring and Harwitz. No one has a definitive answer to this question.

These results reflect a general problem of pure marginal social cost pricing with fluctuating or changing costs. Prices will be less useful in acting as a signal to users and to capital markets if there are significant and frequent fluctuations in costs because there will be a higher variance and hence more uncertainty as to their value at any given time. Lumpy investment is not the only culprit in causing costs to fluctuate. Scope economies and capacity utilization economies also result in costs changing with output or investment levels and hence cause fluctuation in prices based on marginal cost. The prospects of cost recovery with socially efficient pricing depend on how costs change with expansion and contraction of output and capacity. Lumpiness in capacity growth results in fluctuating prices and may lead to over or underfinancing. An assessment must be made of the welfare costs deviating from strict social marginal cost pricing relative to the welfare gains arising from more stable prices which create less uncertainty and have lower transactions cost.

2.4.3 Infrastructure Pricing and the Issue of Equity

Fairness of Efficient Pricing

The economics literature is paying increasing attention to the consequences of selective pricing methods on the equity of various groups of users and non-users; that is, the income distributional consequences of a pricing method. Problems arise when a pricing method, which maximizes economic efficiency, does not necessarily yield what is considered a fair or equitable outcome. The most obvious case is the implementation of congestion (or peak-load) pricing for (urban) transportation. While the introduction of congestion tolls on urban roads would improve the efficiency of resource allocation, such tolls make those commuters with less-flexible work schedules

(which some identify as low- to middle-income groups) pay higher congestion tolls while those with flexible work schedules (generally higher-income groups) may be able to avoid the tolls by shifting commuting time to off-peak hours. Peak-period prices for urban transit services have similar unfavourable impacts on income distribution.

Small (1983) investigated this issue and estimated the net effects of congestion tolls on income distribution. He found that in the absence of any redistribution of toll revenue, low-income groups lose more than high-income groups because the higher value of time savings by high-income groups more than compensates for the toll payment. However, when the way in which the toll revenues are spent is taken into account the distributional impact of toll charges changes. Small examined the income distribution effects of distributing toll revenues to reduce taxes, to subsidize transit or to be used in some other way. The results varied but the basic conclusion was that it is misleading to characterize a congestion or peak-load pricing policy as "regressive."

Peak-period pricing, deemed necessary to achieve the goal of economic efficiency, has been viewed as having negative income distributional effects among various users. Winston (1991) argued that congestion and peak pricing could benefit all income classes if the toll revenues were used to do any of the following: lower property taxes, invest in public transit or replace registration fees or fuel taxes. Therefore, as pointed out by Foster (1974, 1975), the effects of an infrastructure price, such as for a road, runway, or terminal, on income distribution depend on how the collecting agency uses the toll revenue. Congestion tolls and peak-period pricing schemes can be designed to improve both economic efficiency and equity among various groups.

Income Distribution Consequences of Ramsey Pricing

Ramsey pricing allocates fixed or common cost on the basis of willingness to pay. This willingness is a measure of the value to the user and is measured by the price elasticity of demand. This pricing principle has been recommended in circumstances where full or partial cost recovery is desired for underutilized airports, roads and port facilities with economies of traffic density and/or economies of scale (see Morrison, 1982).

The Ramsey principle is used to change prices from marginal (social) cost in such a way as to minimize the efficiency loss associated with departing

from first-best pricing. Ramsey pricing does involve price discrimination since it places a greater fiscal burden on those demanders with less-elastic demands. Some would argue that this is unfair and constitutes an inequity. Economists have no comparative advantage over others in establishing what is or is not fair. One can, however, make the following argument. Having everyone pay exactly the same price is no more fair or equitable than charging different prices. If users are charged the cost for which they are responsible, this seems quite equitable. Different cost responsibility would justify different prices. Furthermore, if those paying the higher prices are as well off or better off, measured in terms of economic welfare, as a result of allowing those paying the lower prices to participate in the market, there is also no inequity.

Fairness in Financing Capacity

The issue of generating funds for capacity expansion raises the question of fairness (equity). At first glance, the generation of funds to cover the financing of capacity through average cost pricing, a common method, appears to be a fair method since each user pays an equal amount. In fact, this is an inequitable (as well as inefficient) method as it entails a subsidy to peak-period users from off-peak users. Short-run, marginal cost pricing in which peak-period prices are set higher than off-peak prices is a more equitable method of raising funds to finance capacity.

Fairness also requires that both fixed and variable costs be allocated to those users responsible for them. Some parts of infrastructure are designed for a sub-group of users, and it is unfair to spread this cost across all users except to the extent that benefits which may accrue to other users, directly or indirectly, may be considered in a Ramsey-type pricing scheme.

The question of financing capacity from users versus the general revenue fund (all taxpayers) is an issue of both efficiency and equity. If marginal cost is less than average cost, efficient pricing will result in a deficit. This deficit must be covered. A partial equilibrium approach to this problem would support a policy of funding the deficit out of general revenue while a general equilibrium approach would argue that the relative welfare losses of funding the deficit users versus the general taxpayer should be considered. As already stated, recent evidence (Jorgenson, 1992; Jorgenson and Yun, 1990; Ballard, Shoven and Whalley, 1985) has shown that the marginal cost of public funds is between \$1.33 and \$1.45. This means that between 33¢ and 45¢ per dollar are lost through a private-sector efficiency loss. It is, therefore,

not obvious that proposing a policy of marginal social cost pricing and ignoring the costs of public funds to finance a deficit will necessarily improve economic welfare. A second-best pricing scheme in which there is a cost-recovery constraint may lead to a higher level of economic efficiency.

2.5 APPLICATIONS OF PRICING PRINCIPLES

Although economists have long argued for the introduction of efficient pricing of transportation infrastructure, little progress has been made by governments. This has been due in part to the view by governments (and of the general public) that infrastructure should be provided at no or a very low charge and that investment in roads and airports promotes economic development and, therefore, excess capacity provides a public good. As such, the allocative and productive efficiency with which infrastructure is provided and operated has not been of concern in most countries. Another important reason for the lack of efficient pricing schemes is the failure by the analysts to provide visible and measurable benefits of efficient pricing.

Until recently, the instances of significant capacity shortfall have occurred only at limited times of the day and at a limited number of locations. However, the recent growth in air travel and the reorganization of the aviation industry together with a failure to invest and price efficiently have led to significant shortages of both runway slots and gates. During the last two decades or so, severe congestion has also occurred on many urban and some intercity roadways. Policy makers could, therefore, rightly ask the questions: What happens if we do implement efficient road or runway pricing? Who gains? Who loses? What is the overall net benefit? These are legitimate questions, and economists have sought to provide answers in a growing applied transportation pricing literature. The outcome of the analysis can be generally divided into two areas: the effect on the use of infrastructure and the impact on investment in infrastructure.

There are few examples of governments introducing prices to achieve the efficient use of roads or runways, even in cases of severe congestion. The introduction of road pricing in Singapore (Holland and Watson, 1978) and the experiment in Hong Kong (Harrison, 1986) are two examples frequently cited. More recently, road pricing has been introduced in Norway (Larsen, 1988). Bergen, Oslo and Trondheim charge fees for vehicles entering a specified area of the city. Oslo and Trondheim use manual and electronic collection systems respectively while Bergen's is a manual system. In

Britain, peak-load pricing has been in effect since the early 1970s to ration airport terminal and runway capacity at London's Heathrow Airport. It was extended to other airports including Gatwick and Stansted in the early 1980s to encourage efficient use of the facilities. The peak and off-peak fees (the combined landing and passenger fees) at the three London area airports are illustrated in Table 2.1. It is important to note from the table that the peak-period fee at Heathrow Airport is 2.5 to 4 times the off-peak fee, depending on the type of aircraft.¹¹

Table 2.1

LANDING FEES AT THE BRITISH AIRPORTS AUTHORITY'S (BAA PLC.) LONDON AREA AIRPORTS
(ALL FIGURES ARE IN UK £, 1991)

Airport	Peak	Off-peak	Weighted average
Heathrow			
B 757 ^a	1680	658	844
Shorts 360	654	153	318
B 747	6259	1747	3336
Gatwick			
B 757	1122	450	634
Shorts 360	444	111	211
B 747	4866	1867	2709
Stansted			
B 757	734	365	507
Shorts 360	123	71	88
B 747	3807	1806	2810

Source: Arthur Reed, "The Unlocking of Heathrow," *Air Transport World* (September 1991), pp. 28–33.

^a The Boeing 757 seats 140 passengers and is a Stage III aircraft, and thus it has a lower noise charge. The Shorts 360 seats 22 passengers. The Boeing 747 seats 270 passengers and is a Stage II aircraft which carries a higher noise charge.

There have been numerous simulation studies using artificial data which have attempted to determine the impact of introducing infrastructure pricing. What would happen to the use of the facility by different user-groups? Would the infrastructure be self-financing? What would happen over the longer term if additional capacity were needed? What is gained (or lost) in undertaking such a policy? These studies use information on existing demand and costs in a modelling framework to address the questions just raised.

Levine (1969) was one of the first to recommend efficient pricing of airport runways. He argued that not pricing efficiently had led to shortages, the excessive peaking of traffic and hence waste of resources such as labour, fuel and time. Carlin and Park (1970), Borins (1978, 1982), Likens (1976), Morrison (1982, 1983, 1987), Gillen, Oum and Tretheway (1988), Morrison and Winston (1989), and Oum and Zhang (1990) all examined various aspects of non-optimal pricing and investments at airports. Newbery (1988), Lee (1982) and Small, Winston and Evans (1989) undertook similar studies but for roadways.¹²

The conclusions of these studies, with their different time periods and data bases, form a consensus in several areas.¹³ First, the existing fee structure has led to a misallocation of traffic and hence the inefficient use of infrastructure (airport and roadway) over the day. For airports, the fee structure, based on aircraft weight, is reasonably consistent with net benefit maximization at airports with sufficient or excess capacity. It is inefficient, however, at congested facilities (Morrison, 1987). The existing fee structure benefits some groups over others. In particular, general aviation, commuter and regional (local) carriers are beneficiaries while international and trunk carriers are losers (Morrison, 1983, 1987). Similar results hold for roadways. Newbery (1988) noted that vehicles impose four costs on society: road damage costs, congestion costs, accident externalities and environmental pollution costs. The current system of licence fees and fuel taxes has, as with the case of airports, resulted in a misallocation of traffic over the day and a misallocation between types of vehicles. Rural automobiles are charged too much while urban-peak automobiles and large four-axle trucks on a rural road pay too little.

Lee (1982) provides a set of efficient user-charges on the U.S. intercity roads for a sample of vehicles.¹⁴ These are reported in Table 2.2. Gillen, Oum and Tretheway (1988) have undertaken a similar type of analysis but for airport infrastructure (Pearson International Airport, Toronto). Their results are reported in Table 2.3. These two studies illustrate the magnitude of the current underpricing to different user-groups and the magnitudes of the changes anticipated under socially efficient pricing. Which user-classes are gaining under the present system and which are losing is also evident from these studies. Furthermore, this illustration provides an intuitive feel for the notion that the current pricing system creates inefficiencies from the choice of vehicle type and the time of travel.

Table 2.2

EFFICIENT USER-CHARGES ON THE U.S. ROADS FOR A SAMPLE OF VEHICLES
 (VALUES EXPRESSED IN CENTS/VEHICLE-MILE OF TRAVEL IN 1981 U.S. DOLLARS)

Vehicle type	Location	Key parameter	Pavement repair	User-costs	Admin.	Excess delay	Air pollution	Noise	Total	Existing average user-fee
Car (3000 lb)	rural	v/c=.05			0.3	0.3			0.6	1.3
Car (3000 lb)	urban	v/c=.85			0.7	11.2	1.5	0.1	13.5	1.7
Single unit, 3 axle truck (40,000 lb GVW)	small urban	v/c=.35 PCE=1.2 ESAL=.8	25.6	7.5	0.5	2.2	0.2	0.2	36.2	4.8
Combination truck, 5 axle (72,000 lb GVW)	rural interstate	v/c=.15 PCE=1.2 ESAL=1.6	8.0	5.9	0.3	0.4			14.6	9.0
Combination truck, 5 axle (72,000 lb GVW)	urban interstate	v/c=.35 PCE=1.2 ESAL=1.6	24.0	16.3	0.3	1.4	3.0	4.0	49.0	9.0
Combination truck, 4 axle (100,000 lb GVW)	rural arterial	v/c=.05 PCE=3.0 ESAL=27.2	408.0	95.2	0.3	0.3		0.2	504.0	5.0

Source: Reproduced from Lee (1982).

Notes: v/c — volume/capacity; PCE — Passenger Car Equivalent; ESAL — Equivalent Standard Axle Load

Table 2.3

**COMPARISONS BETWEEN SOCIAL MARGINAL COSTS AND 1985 LANDING FEES FOR SELECTED AIRCRAFT
PEARSON INTERNATIONAL AIRPORT, TORONTO
(1988 CAN \$)**

	Aircraft type					
	B747-200	DC10-40	B737-200	Dash 8	Business jet	Piston
1985 fees						
Domestic	\$521	\$355	\$ 80	\$ 12	N/A	\$ 0
International	\$769	\$524	\$ 85	\$ 15	N/A	\$ 0
Social marginal cost						
<i>High months</i>						
High-peak	\$426	\$376	\$269	\$213	\$211	\$161
Low-peak	\$246	\$196	\$154	\$ 98	\$105	\$ 55
<i>Low months</i>						
High-peak	\$271	\$221	\$170	\$114	\$120	\$ 70
Low-peak	\$226	\$176	\$142	\$ 86	\$ 93	\$ 43
Off-Peak	\$206	\$156	\$129	\$ 73	\$ 81	\$ 31

Source: Gillen, Oum and Tretheway (1988).

Gillen, Oum and Tretheway (1988) investigated efficient pricing at Pearson International Airport. They developed a set of socially optimal peak-period prices based on measures of variable costs with runway use, measures of congestion externalities, noise externalities and fixed costs for airfield operations. A comparison of the calculated social marginal costs and the current landing fee for selected aircraft are described below. They pointed out that the fee structure used by Transport Canada is common across all Group I airports regardless of differences in available capacity and demand variation. Furthermore, the fees are weight-based and discriminate against non-domestic flights. Both generate inefficiencies.¹⁵

The implementation of this pricing structure was simulated to reduce annual light-plane movements from 63,000 to 28,000 with a proportionately higher reduction in the peak periods. The social marginal cost pricing of runways (noise externalities and congestion were included) would have generated approximately \$50 million instead of the actual airfield revenue of \$26.6 million in fiscal 1985-86. They also developed a set of terminal

user-charges which varied with the peak. These were calculated to be \$2.82 and \$6.49 per passenger for peak domestic and international flights respectively and \$1.54 and \$3.54 in the off-peak. (The current charge is \$1.00 per seat for domestic flights and \$2.30 per seat for international flights.) The gainers and losers are the same as those identified by Morrison.

The second area of consensus in the applied pricing literature is the financing of infrastructure. Under the status quo, there is a significant deficit for roadways and to a lesser extent for airports and airways in a number of jurisdictions. The move to socially efficient or efficient second-best pricing improves the financial condition of the infrastructure agency. If social marginal cost pricing is applied to an uncongested facility, there will be a deficit. Thus, rural roads or uncongested airports would require a second-best pricing scheme to break even or a subsidy if marginal cost pricing is applied. For uncongested airports, for example, Morrison (1982) illustrated the set of Ramsey prices which yield the maximum social benefits subject to revenue covering all costs. On the other hand, a congested facility is likely to be self-financing. Oum and Zhang (1990) provided a useful example for airports. A corresponding example for roads was provided by Small, Winston and Evans (1989).

The third area in which there is some consensus is that the benefits from introducing social marginal cost pricing are significant. Borins (1982) reported that efficiency losses due to non-optimal pricing were significant for Pearson International Airport and also for the access roads to the airport. Losses were measured in terms of economic surplus defined as the difference between what individuals are willing to pay for transportation services and the social costs of providing those services. Small, Winston and Evans (1989) showed that efficient pricing of all U.S. roads with the current investment would yield an annual net welfare gain of \$5.4 billion (1988 dollars), and if efficient investment policy was coupled with efficient pricing policy, the gain would increase to \$7.75 billion annually.¹⁶ Most of the gain would come from savings in pavement costs (maintenance and capital costs) which were reduced from \$20 billion to \$13 billion in 1982. The sources and distribution of the efficiency gains are listed in Table 2.4.

Morrison and Winston (1989) provided similar results for all U.S. airports. The welfare gain from efficient pricing would be \$3.8 billion annually. Efficient pricing in conjunction with efficient investment would raise this

amount to \$11 billion annually. This is relative to the status quo. The values varied with assumptions regarding the elasticity of demand and value of time. The values reported here are conservative values for both demand and time. The distribution of the benefits is reported in Table 2.5.

Table 2.4
ECONOMIC EFFECTS OF EFFICIENT INFRASTRUCTURE PRICING FOR THE U.S. HIGHWAYS
 (IN BILLIONS OF 1982 U.S. DOLLARS)

	Efficient pricing and investment (first-best pricing)	Efficient pricing with current investment	Efficient investment with current pricing
Investment costs			
Maintenance savings	9.426	6.441	8.536
Capital savings	-1.276	0.0	-2.236
Total savings	8.152	6.441	6.300
Trucking firms' and shippers' gains	0.134	-5.586	0.0
Gains from modal shifts	0.040	0.615	0.0
Government revenue	-0.574	3.884	-5.190
Total benefits	7.752	5.354	1.110

Source: Small, Winston and Evans (1989).

Table 2.5
ANNUAL ECONOMIC EFFECTS OF EFFICIENT INFRASTRUCTURE POLICY FOR AIRPORTS
 (CHANGE RELATIVE TO CURRENT PRACTICE, IN BILLIONS OF 1988 U.S. DOLLARS)

Item	Efficient pricing and runway investment	Efficient pricing with current runway investment
Consumer surplus change from landing and takeoff fees	1.10	-12.53
Reduced delay to travellers	7.91	3.62
Carriers' operating cost savings	2.77	1.23
Airport revenues less costs	-0.77	11.50
Total welfare change	11.01	3.82

Source: Winston (1991).

The airport pricing literature makes it quite clear that the failure to introduce landing fees, which more closely reflect the real costs of usage and externalities of the facilities, has generated significant economic waste. This waste is measured in real resource costs to carriers who use crew time, fuel and capital in congested facilities.¹⁷ It is also measured in wasted time to travellers. These delay costs can mount into the billions of dollars.

2.6 SUMMARY

There is a broad literature which has examined almost every aspect of infrastructure pricing from simple congestion models with homogeneous traffic to infrastructure pricing with several user-groups, uncertainty, varying demands and intermodal congestion. Despite the varied approaches and level of complexity in the models, all studies find that optimal prices for peak-period use are well above those currently charged. Calculated net benefits from measures which introduced some “demand management” were significant. This applies to all modes of transportation in the intercity passenger system despite differences between modes.

The motivation to develop socially efficient prices arises from the need to limit the use of existing facilities and to maximize economic efficiency. The latter objective is not some idol but rather is based on the argument that such pricing would make society better off by having both the efficient use of a facility and an optimal amount invested in infrastructure. Gillen, Oum and Tretheway (1988), for example, have shown that the current landing fee structure in Canada (and elsewhere), which is based on the weight of the aircraft, is inefficient in a situation where congestion exists. Indeed, they demonstrate that the largest proportion of marginal social costs is made up of movement delays to other aircraft and noise externalities. Thus, current landing fees, which ignore the congestion externality, undercharge aircraft in inverse proportion to their weight.

The air transport sector is subjected to problems not unlike those found in urban areas — peak-period demands with multiple user-groups. The road and airport runway pricing literature demonstrate that an effective solution to the problems of airport congestion is efficient pricing *and* investment policies. This is an important point. In the past, supply management was the single strategy. Some economists, as well as other transportation experts,

have focussed almost entirely on demand management through pricing. This study argues that it takes a combination of the two, pricing and investment, to yield an efficient long-run outcome.

The basic message is quite clear. Continuing the supply-side approach to solving the economy's transportation problems is not only difficult in a situation of fiscal restraint, it foregoes economic gains which are available independently of whether fiscal restraint applies. Demand management is part of an efficient approach to infrastructure management and involves setting user-charges for roadway, airport or waterway infrastructure to reflect their social costs.

Such a pricing scheme has three significant impacts. First, economic welfare increases as the demand for infrastructure is rationed more efficiently in a way which reflects the costs of using the infrastructure. Second, intermodal distortion is reduced, even with a second-best Ramsey pricing. Under the current regime of infrastructure fees the distribution of traffic among modes can be expected to result in an efficiency loss. Finally, the current pricing policy results in some infrastructure being financed out of general tax revenues. This is not necessarily inefficient if the community receives benefits as a result of the infrastructure (if there are positive externalities sufficient to justify such a subsidy). The current system also ensures that the recurring infrastructure problems of the past will continue into the future.¹⁸ For congestion-prone facilities, basing infrastructure prices on social costs would generate sufficient funds to provide an optimal facility and have the infrastructure more likely to be self-financed.

Winston (1991) summed it up best when he said:

Surprisingly, the belief of most economists that public infrastructure spending should be substantially increased is not based on efficient pricing and investment principles. Instead, it appears to be based on either personal observations or on a suspicion that because uncongested infrastructure is a public good, society has tended to invest too little in it. Both perspectives have diverted many economists and policymakers from realizing there are surprisingly large but plausible benefits from *efficient* infrastructure pricing and investment. (p. 114)

Marginal social cost pricing is also not incompatible with fairness or equity. Congestion or peak-load pricing in conjunction with a strategy to use the revenues can generate net positive benefits for society. Marginal social cost pricing corrects distortions rather than introduces them. Some groups are made better off and others worse off but this should not justify rejection of this pricing policy. As Hau (1991) noted, it is perhaps asking too much of a pricing mechanism to solve the pricing, investment *and* income distribution problems. One way of viewing this is to use road and airport funds from efficient pricing to both invest efficiently in capacity (new roads or runways) and to use any residual to satisfy the principles of equity taxation. Thus, a transportation fund could potentially invest in public transportation, rural roads or small airports. The magnitude of the fund would depend on scale economies and indivisibilities.

Social marginal cost pricing does not necessarily yield a surplus or break even. When the total revenue from social marginal cost pricing is not enough to cover total cost, there are three options open to the infrastructure authority:

- to continue to use social marginal cost pricing with subsidization from the general tax revenue;
- to use Ramsey pricing which provides a second-best solution with a break-even financial performance; or
- to adopt two-part pricing (access/usage tariffs).

The Ramsey pricing minimizes the loss of economic efficiency caused by deviating prices from the respective marginal costs in order to achieve financial break even. In effect, it charges a higher markup to the less price-elastic product or market segment by setting the markup in inverse proportion to the price elasticity demand.

Two-part pricing consists of a flat fee for the right to access a facility (for example, vehicle licence fee to access the road system), and a usage fee (for example, charge per kilometre of road usage). Two-part pricing leads to first-best optimality if the demand for access is not price sensitive, and if usage charges are set at the marginal costs of usage and the access fee is set at a sufficiently high level to allow the firm to break even. Furthermore, under these conditions, the regulation of the access fee alone can induce the monopoly firm to charge marginal cost as the usage price. When demand

is sensitive to price, the two-part tariffs are computed by applying the Ramsey rule to the access and usage demands as if they were two separate products with interrelated demands. This then becomes a form of second-best pricing.

The use of first-best pricing with subsidies from general revenues has been long advocated in the transportation literature. There is increasing evidence, however, that the efficiency costs of raising revenue through taxes may more than offset the relative efficiency costs of moving from first- to second-best pricing schemes.

3. CARRIER AND INFRASTRUCTURE COST STRUCTURES: A LITERATURE SUMMARY

3.0 INTRODUCTION

This section summarizes the theoretical and empirical literature on the cost structure of modal services (carriers) and of the provision of infrastructure. Since the infrastructure planner must establish user-charges and make capacity investment decisions to maximize the economic welfare of society, understanding the behaviour of the combined cost of a carrier's service provision and infrastructure provision is essential for the development of a set of socially optimal prices for infrastructure.¹⁹ If short-run costs fall because of increased capacity utilization but long-run costs exhibit constant returns to scale, it is still possible to have marginal social cost pricing and fully cover costs. If, however, long-run costs are characterized by some economies, a second-best pricing approach will be required to have total revenues cover costs and minimize the efficiency loss of deviating from first-best pricing. A prerequisite to understanding the structure of the combined cost is to understand each component, that is, carrier cost structure and infrastructure cost structure.

Social benefit maximization is the objective of infrastructure pricing and investment. The relevant cost is, therefore, the full social cost a user of transportation services imposes on society as a whole. The full social cost includes not only all private resource costs such as payments to labour, capital and energy but also external costs the user imposes on others, such as congestion costs and noise and air pollution costs. The remainder of subsection 3.0 describes several cost concepts which will be used in the

sections following. Subsection 3.1 summarizes the literature on several key aspects (economies of scale, economies of traffic density and economies of scope) of the carriers' cost structure for each mode. The empirical and theoretical literature on the cost structure in the provision of infrastructure services is presented in subsection 3.2. Finally, subsection 3.3 discusses the implications of the empirical results for infrastructure pricing.

External Costs and Internal Costs

Internal costs, sometimes referred to as private costs, are the costs borne by the supplying agency. These are normally, but not always, financial costs incurred as a result of purchasing factor services in the market and would include labour wages, interest on capital and the price of fuel. Although external costs represent genuine resource costs, they do not directly influence decisions of transport suppliers in their provision of transport services. The external costs include general social and environmental impacts and the cost of congestion delay imposed on third parties. It also includes the effects on non-users. There is a clear distinction between internal costs that influence the optimal choice of the transportation supplier and the external costs which affect others, but do not directly influence the supplier's decisions. However, both must be considered in the socially optimal design and pricing of infrastructure. For example, in selecting the frequency and timing of flights and type of equipment use at an airport, an airline normally does not consider the noise and air pollution resulting from such decisions.

Short-Run Versus Long-Run Costs

Long-run costs, using the standard economic definition, are all variable; there are no fixed costs. In the short run, the ability to vary costs in response to changing output levels and mixes differs among the various modes of transportation. Since some inputs are fixed, short-run average cost is likely to continue to fall as more output is produced until full capacity utilization is reached. Economies of traffic density are another potential source of cost economies in transportation. Unit cost per passenger-kilometre decreases as traffic volume increases over a fixed network. The fixed network means this is a short-run rather than long-run cost. Density economies are a result of using a network more efficiently. The potential for density economies depends on the configuration of the network. Carriers in some modes, such as air, have reorganized their network, in part, to realize these economies.

In the long run, additional investment is needed to increase capacity and/or other fixed inputs. The long-run average cost curve, however, is formed by the envelope of the short-run average cost curves. For some industries, the long-run average cost often decreases over a broad range of output as firm size (both output and capacity) expands. This is called "increasing returns to scale" or "economies of scale." The presence of economies at the relevant range of firm size means that the larger the size of the transportation firm, the lower the per-unit cost of output. These economies of scale may potentially take a variety of forms in transportation services and may vary significantly according to the mode of transportation involved.

Common and Joint Costs

The production of transport services in most modes involves joint and common costs. A joint cost occurs when the production of one good inevitably results in the production of another good in some fixed proportion. For example, consider a rail line running only from point A to point B. The movement of a train from A to B will result in a return movement from B to A. Since the trip from A to B inevitably results in return trip costs, joint costs occur. Some of the costs are not traceable to the production of a specific trip so it is not possible to fully allocate all costs nor to identify separate marginal costs for each of the joint products. For example, it is not possible to identify a marginal cost for an i to j trip and a separate marginal cost for a j to i trip. Only the marginal cost of the round trip is identifiable.

Common costs occur when the facilities used to produce one transport service are also used to produce other transport services. For example, track or terminals used to produce freight services are also used for passenger services. The production of a unit of freight transportation does not, however, automatically lead to the production of passenger services. Thus, unlike joint costs, the use of transport facilities to produce one good does not inevitably lead to the production of some other transport service since output proportions can be varied. The question arises whether or not the presence of joint and common costs could prevent the market mechanism from generating efficient prices. A substantial literature in transport economics (Mohring, 1976; Button, 1982; Kahn, 1970) has clearly shown that conditions of joint, common or non-allocatable costs do not preclude economically efficient pricing.

3.1 CARRIER COSTS

There are three key indicators of the cost characteristics of a firm. They include economies of scale and economies of scope which are long-run concepts. The other is economies of density. How do they influence costs? Why are they important to this discussion of transport infrastructure pricing? These questions will be addressed in the following subsections.

Economies of Scale

There has been some confusion in the literature between economies of scale and economies of density. The distinction is important since scale is a long-run concept and thus affects industry structure whereas density is a short-run concept and is more meaningful for industry behaviour such as pricing. Density economies are said to exist when a 1 percent increase in all outputs (holding network size, production technology and input prices constant) increases the firm's cost by less than 1 percent. In contrast, scale economies exist when a 1 percent increase in output²⁰ and size of network increases the cost by less than 1 percent, with production technology and input prices held constant.

Economies of scale refer to a long-run average cost curve which slopes down as the scale of the transport firm increases. The presence of economies of scale means that as the size of the transport firm gets larger, the average or unit cost gets smaller. Since most industries have variable returns to scale cost characteristics, whether or not a particular firm enjoys increasing, constant or decreasing returns to scale depends on the overall market size and the organization of the industry.

The presence or absence of scale economies is important for the industrial structure of the mode. The presence of significant scale economies implies that fewer larger carriers would be more efficient and this, under competitive market circumstances, would naturally evolve over time. Scale economies are important for pricing purposes since the greater the scale economies, the more average and marginal costs deviate. It would, therefore, be impossible to avoid a deficit from long-run marginal (social) cost pricing.²¹

Economies of Traffic Density

Although they have a different basis than scale economies, economies of density can also contribute to the shape of the modal industry structure and

affect the way a carrier organizes the delivery of its service, spatially. The magnitude of density economies also depends on the size of the market. In the air market, for example, deregulation has allowed carriers to respond to market forces and obtain the available density economies to varying degrees. Canadian carriers have been less successful than their counterparts in the United States in realizing the maximum density economies because the market is smaller in Canada.

Keeler (1974), Harris (1977), Friedlaender and Spady (1981) and Levin (1981) have all shown that there are large increasing returns to traffic density in the U.S. railroad industry. They show that allowing all factors of production to vary, including information but not route mileage, a railway producing 10 million revenue tonne-miles per mile of road, for example, will have substantially lower average costs than a railway producing only 5 million revenue tonne-miles per mile of road. Harris (1977) estimated that approximately one third of density economies were due to declining average capital costs, and two thirds due to declining fixed operating costs, such as maintenance and administration.

Gillen, Oum, and Tretheway (1985, 1990), using data from the airline industry in Canada, illustrated that unit costs would decrease for all carriers, except Air Canada, if they carried more traffic within their given network. The authors also noted that a large portion of the density economies resulted from fixed costs associated with a network — costs which are independent of level of output.

Economies of Scope

Typically, the transport firm produces a large number of conceptually distinct products from a common production facility. In addition, the products of most transportation carriers are differentiated by time, space and quality. Because a number of distinct non-homogeneous outputs are being produced from a common production facility, joint and common costs occur. The presence of joint and common costs gives rise to economies of scope. There has been some confusion in the multi-product literature among the concepts of sub-additivity of the cost function, trans-ray convexity, inter-product complementarity and economies of scope. Sub-additivity is the most general concept and refers to a cost function which exhibits the characteristic that it is less costly to produce any amounts of any number of goods in one plant or firm than to subdivide the products or service in any

proportion among two or more plants. Trans-ray convexity is a somewhat narrower concept. It refers to a cost function which exhibits the characteristic that for any *given set* of output vectors, the costs of producing a weighted average of the given output vectors is no greater than the weighted average of producing them on a stand-alone basis. Economies of scope refers to the cost characteristic that a single-firm, multi-product technology is less costly than a single-product, multi-firm technology. It, therefore, is addressing the issue of the cost of adding another *product* to the product line. Interproduct complementarity is a weak test of scope economies. It refers to the effect on the marginal cost of one product when the output of some other product changes. It, therefore, is changing the *amount of output* of two or more products and not the *number of products*. Whether scope economies exist, and the extent to which they exist, depends on both the number of products and the level of each output. No reliable empirical estimates of economies of scope for transportation modes exist which are both based on reliable data and undertaken in a theoretically consistent fashion.

3.1.1 Air Carriers

A considerable number of studies, Douglas and Miller (1974), Keeler (1974), Caves, Christensen and Tretheway (1984), Caves, Christensen, Tretheway and Windle (1985), McShan and Windle (1989) and Gillen, Oum and Tretheway (1985, 1990), have been directed at determining the functional relationship between total per-unit operating costs and firm size in airlines. All studies have shown that returns to scale are roughly constant; thus, size does not generate lower per-unit costs. Generally, however, the measures of economies of density illustrate that unit cost would decrease for most carriers, in both Canada and the U.S., if they carried more traffic within their given network.²² In other words, the industry experienced increasing returns to density. The results also indicated that the unexploited economies of density are larger for low-density carriers.

Caves, Christensen, and Tretheway (1984) have shown that it is important when measuring costs to include a network size variable in the cost function, along with output, which would allow for the distinction between returns to scale (RTS) and returns to density (RTD). McShan and Windle (1989) used the same data set as that used by Caves et al., and explicitly account for the hub-and-spoke configuration that has evolved in the United States since deregulation in 1978. They estimated a long-run cost function which employs all the variables included in Caves et al., and found the

returns to density of about 1.35. The hubbing variable indicates that, *ceteris paribus*, a carrier with even 1 percent more of its traffic handled at hub airports expects to enjoy 0.11 percent lower cost than other similar carriers.

3.1.2 Intercity Buses

According to Gillen and Oum (1984), the hypothesis of constant returns to scale could be rejected for the intercity bus industry in Canada where they found decreasing returns to scale at the mean of the sample (0.91). Large firms exhibited strong decreasing returns to scale, and small- and medium-sized firms exhibited slight departures from constant returns. These empirical measures may be biased, however, since no measure of network size and market density was included in the estimation. This exclusion would have a differential impact on the measure of scale economies depending on the route mix of each firm. No cost complementarities were found to exist between the three outputs, namely, number of scheduled passengers, revenue vehicle miles of charter, tour and contract services and real revenue from freight. These results, however, are biased since no network measure was included in the estimating equations. The scale economy measure, therefore, contains some of the influence of available density economies.

Since deregulation of the intercity bus industries in the United States and the United Kingdom, the number of firms has been reduced significantly. In the absence of scale economies, the forces leading to this industry structure would include density economies. Route reorganization, for example, has been observed to approximate hub-and-spoke systems and the use of smaller feeder buses on some rural routes.

The industry reorganization is similar to what occurred in the airline industry. The consolidation of firms was driven by density and not scale economies. One significant difference between these two industries, however, is that airline demand has been growing while intercity bus demand is declining.

3.1.3 Trucking

Several empirical analyses of the trucking production function have shown that the long-run marginal cost and average cost curves are relatively flat with respect to changes in the level of output. In other words, economies of scale are either absent or very small. Koenker (1977) showed that there are very small economies of scale present up to relatively low levels of output

and, thereafter, unit costs tend to rise gradually. Friedlaender and Spady (1981) also found mild diseconomies of scale in the production of trucking services as well as cost complementarities in the production of the multiple outputs trucking firms produce. Friedlaender and Bruce (1985) found that, for the period from 1974 to 1979, larger firms experienced greater diseconomies of scale than smaller firms. In 1979, however, the situation was reversed and, in their study, the authors suggested that the larger carriers were reaping the benefits from longer hauls.

Friedlaender and Chiang (1984) examined the effect of various network variables on trucking costs. Their study found that at the mean of the data there were constant returns to scale, while, as with other modes, there were significant gains from better network utilization, hence economies of density. Ying (1990) examined the impact of deregulation on the productivity in the United States Class I and II trucking industry. He found that at the sample mean, a 1 percent increase in output caused total cost to increase by about 1.073 percent, suggesting very slight decreasing returns to scale. It is unlikely, however, that given the size of the standard error this result is statistically significantly different from 1.00 or constant returns to scale.

3.1.4 Railways

The structure of railway costs is generally characterized by high fixed costs and low variable costs per unit of output. The essential production facilities in the railway industry exhibit a significant degree of indivisibility. As with other modes, the production of railway services gives rise to economies of scope over some output ranges. For example, track and terminals used to produce freight services are also used to produce passenger services.

Caves, Christensen and Tretheway (1980) found that the United States railway industry was characterized by constant returns to scale over the relevant range of outputs. However, their sample did not include relatively small railroads, firms with less than 500 miles of track. Griliches (1972) and Charney, Sidhu and Due (1977) found increasing returns to scale for such small U.S. railroads. Friedlaender and Spady (1981) suggested that there may be very small economies of scale with respect to firm size. Keeler (1974), Harris (1977), Friedlaender and Spady (1981) and Levin (1981) all found large and significant economies of traffic density in railway services. Friedlaender and Spady (1981) estimated a short-run cost function with five variable inputs,

one quasi-fixed factor (structures) and two outputs which took the form of hedonic functions, accounting for factors such as low-density route miles and traffic mixes. The study found constant returns to scale but increasing returns to traffic density. Caves, Christensen, Tretheway and Windle (1985) examined economies of scale and density in the United States railroads. Their basic result demonstrates that there are substantial returns to density in the U.S. railway operations.

The economies of traffic density and economies of scale estimated by various studies are compared in Table 3.1.

Table 3.1
RETURNS TO DENSITY AND SCALE IN U.S. RAILWAYS

Study	Density	Scale
Friedlaender and Spady (1981)	1.16	.88–1.08
Caves, Christensen and Swanson (1981)	—	1.01
Harmatuck (1979)	1.92	0.93
Harris (1977)	1.72	1.03
Keeler (1974)	1.79	1.01
Caves et al. (1985)	1.76	0.98

Source: Table 4.3 in Caves, Christensen, Tretheway and Windle (1985).

Note: Estimates of returns to scale are for fixed length of haul and trip length.

3.2 INFRASTRUCTURE COSTS

As early as 1962, Mohring and Harwitz demonstrated that the financial viability of an infrastructure facility, under optimal pricing and investment, depended largely on the characteristics of its cost function. To quote Winston (1991):

If capacity and durability costs are jointly characterized by constant returns to scale, then the facility's revenue from marginal cost pricing will fully cover its capital and operating costs. If costs are characterized by increasing returns to scale, then marginal cost pricing will not cover costs; conversely, if costs are characterized by decreasing returns to scale, marginal cost pricing will provide excess revenue.
(p. 115)

The objective of this subsection is to summarize the theoretical and empirical literature on the cost characteristics of modal infrastructure. The discussion deals with the following types of infrastructure: airports, highways and railways.

In developing a set of socially efficient prices for modes of intercity transport, it is not just the carrier's cost structure which is important. Airports, roadways and harbours all represent public capital which is used by the carriers in the different modes to produce and deliver their modal services. This capital must also be priced in an efficient way to achieve the economic welfare gains available from economically efficient pricing. As with the carriers, the ability to apply first-best pricing principles to infrastructure and still satisfy cost-recovery constraints depends on the cost characteristics of building and maintaining the infrastructure.

Cost characteristics include scale economies, scope economies, density economies and utilization economies. Scale economies refer to the size of a facility. For example, is it cheaper per unit to build three runways than it is to provide two runways? If so, there are economies of scale in the provision of runways. Scope economies encompass similar concepts as with carriers. Small, Winston and Evans (1989) referred to scope economies in highways when both volume and durability are supplied. Volume refers to the number of lanes while durability refers to the ability to carry heavier vehicles. A similar concept would apply to airports (small and large aircraft, VFR and IFR traffic) and harbours (large ships and small ships). Although rail infrastructure is currently supplied by the railways, there have been moves to separate infrastructure and carrier services. This separation would mean pricing the track and terminals separately from carrier services.

Density economies should also, in principle, be evident in the provision of infrastructure. It is, for example, possible to expand outputs and all inputs for highways while holding the size of the network fixed. There are no empirical estimates of these types of economies for any mode.

Utilization economies refer to the short-run cost function. They describe how quickly average and marginal costs fall as capacity utilization approaches capacity. Although not of direct interest, they are important to consider in any cost estimation since failure to consider capacity utilization can bias upward the measures of both long-run average and marginal costs.

3.2.1 Airports

Economists have typically assumed that capacity expansion is divisible. In his analysis of the optimal pricing and investment in airport runways, Morrison (1983) showed that airport capacity construction is characterized by constant returns to scale. Therefore, under perfect divisibility of capacity expansion, the revenue from tolls would be exactly equal to the capital cost of capacity investment (Mohring and Harwitz, 1962). Morrison's results, however, were based on a sample of 22 of the busiest airports in the United States and did not include any small airports. In the literature, there is no empirical evidence on the cost characteristics of capacity construction of new small airports or capacity expansion of existing small airports (for example, one runway).

3.2.2 Highways

In general, highways produce two outputs, traffic volume which requires capacity in terms of the number of lanes and standard axle loadings which require durability in terms of the thickness of the pavement. Before determining economies of scale in this multi-product case, the measure of economies of scale for each output, or the product-specific returns to scale, must be examined. Small, Winston, and Evans (1989) reported the existence of significant returns to scale associated with the durability output of roads and the ability to handle axle loads. This is because the pavement's ability to sustain traffic increases proportionally with its thickness. They also found evidence that there are slight increasing returns to scale in the provision of road capacity; that is, the capacity to handle traffic volume. However, they reported diseconomies of scope from the joint production of durability and capacity because, as the road is made wider to accommodate more traffic, the cost of any additional thickness rises since all the lanes must be built to the same standard of thickness. They concluded that these three factors together result in highway production having approximately constant returns to scale. In other words, the output-specific scale economies are offset by the diseconomies of scope in producing them jointly.

3.2.3 Railways

An important difference between rail and other modes of transportation is that most railroads provide the infrastructure themselves, and the pricing is undertaken jointly for carrier services and infrastructure. In a few cases,

however, ownership and/or management of the trackage has been separated from carriers. Sweden is a good example but even in the United States there have been joint running rights on tracks. This creates a situation where one firm may be responsible for the provision of trackage and another for carrier services. It is, therefore, legitimate to ask if there are any scale economies in the provision of railway infrastructure. There are no empirical estimates but it may be possible to use some of the Small, Winston and Evans (1989) work for roads to shed some light on the issue.

Small et al. argued that road infrastructure produces two outputs, durability and capacity. The former refers to the thickness of roads and the latter to their width. They found increasing returns with respect to durability. This is less likely to occur with a rail line since there would be a relatively broad range of rail car axle loadings for a given level of durability of rail, ballast and ties. Thus, there may only be minor economies. The authors found some minor increasing returns to scale in the provision of capacity. One would expect these same economies would exist for rail since doubling track more than doubles capacity (Gillen and Oum, 1984). Small et al. found diseconomies of scope from the joint production of durability and capacity for highways. These diseconomies are less likely to be evident in rail due to the broad range of durability noted above and the ability to allocate traffic to specific tracks. On balance, there may be generally constant or minor increasing returns in the provision of rail line infrastructure. This conclusion, however, is reached from an intuitive discussion of capacity expansion costs and not from empirical estimates.

The output-specific scale economies seem to be minor as do the diseconomies of producing them jointly. In Canada since VIA Rail leases track from CNR and CPR, the relevant question is, what is the optimal price for VIA's use of tracks?

3.3 SUMMARY OF THE COST STRUCTURE FOR CARRIERS AND INFRASTRUCTURE

The full costs of a mode are the sum of infrastructure costs and modal services costs. Since the choice of a particular basis for infrastructure pricing will influence the modal choices of the end users, optimal pricing strategies and cost recovery should consider the combined cost of infrastructure provision and carrier (or user) costs in order to maximize social welfare. If

markets for carrier services are competitive and there are constant returns in the provision of infrastructure for the mode, marginal (social) cost pricing will yield a socially efficient outcome and full cost recovery. If there are economies, from whatever source in the provision of infrastructure, first-best pricing may result in a deficit requiring a subsidy from general revenues which has consequences for economic welfare. Second-best pricing to recover costs may also lower social welfare. The issue is which minimizes the loss.

3.3.1 Air

A number of studies have been directed at determining the behaviour of an airline's cost function with respect to changes in the level and composition of output. The studies have shown that the long-run average cost curve is relatively constant over a wide range of output associated with different scales of plant; that is, there are no economies of scale in the airline industry. This means that the size of a carrier does not generate lower per-unit costs. Gillen, Oum and Tretheway (1985, 1990), however, found that the airline industry in Canada experienced increasing returns to traffic density; that is, the unit cost would decrease for all carriers (except Air Canada) if they carried more traffic within their given network.

Studies concluded that airport capacity construction is also characterized by constant returns to scale. This implies that the combined cost of carriers and infrastructure is also characterized by constant returns to scale.²³

3.3.2 Road

There are somewhat different results for intercity bus and truck. Several empirical studies of the trucking industry found constant returns to scale in the industry while studies on the intercity bus industry found the hypothesis of constant returns to scale rejected in favour of decreasing returns to scale (0.91). The research also found no economies of scope between the three outputs, namely, scheduled passenger, charter and contract services. There is no empirical evidence on density economies. Observing the parallel mergers which have occurred in the United States and the United Kingdom bus industries after deregulation, however, one might hypothesize that there are density economies.

Road infrastructure yields two outputs, namely, traffic volume which requires capacity (measured in number of lanes) and standard axle loadings which require durability (measured in thickness of pavement). Small, Winston, and

Evans (1989) reported the existence of significant economies of scale with respect to the durability of road and mild returns to scale with respect to traffic volume. However, they reported diseconomies of scope from the production of both durability and traffic volume because, as the road is made wider to accommodate more traffic, the cost of any additional thickness rises since all the lanes must be built to the same standard of thickness. The final outcome of these three factors at work is that highway capacity construction is characterized by approximately constant returns to scale. In other words, the output-specific economies of scale are offset by the diseconomies of scope for having to produce them jointly. Since they included both infrastructure costs and the costs incurred by road users (individual drivers and transportation carriers) in the total cost of highway modes, their result on the overall constant returns to scale is for the combined cost of highways and users.

3.3.3 Rail

Currently, an important difference between the railway mode and other modes is that rail infrastructure is provided by carriers and thus the infrastructure cost is reflected in the freight rates and passenger fares. Since railway companies provide their own infrastructure (VIA Rail in Canada is an exception and Amtrak in the United States is a partial exception) the carrier's cost structure represents those of the combined carrier and infrastructure costs.

Several studies in the United States have shown that the railway industry is characterized by constant returns to scale over the relevant range of output. Studies have indicated, however, that economies of scale are present for small-sized firms. On the other hand, all studies have shown that there are large and significant economies of traffic density in railway services.

4. A FRAMEWORK FOR THE ANALYSIS OF OPTIMAL USER-CHARGES AND COST RECOVERY

4.1 GENERAL FRAMEWORK FOR USER-CHARGES AND COST RECOVERY

In this chapter, the general framework for integrating optimal user-charges and investment is presented and applied to airport and road infrastructure. A similar approach to that proposed by Small, Winston and Evans (1989) is

taken to model the integrated pricing and investment decision. This approach has an advantage of treating the choice of price, capacity and a durability simultaneously. This model can be applied to all types of infrastructure but this study presents the analysis using roads as the example.

The social benefit from infrastructure, defined as the difference between what ultimate consumers are willing to pay and the combined cost of providing modal services and infrastructure, can be maximized by choosing price, capacity and a durability standard which maximizes the following expression:

$$\text{MAX}_{Q_t, W, D} \text{NB} = \sum_{t=1}^T \left(\int_0^{Q_t} P_t(Q) dQ - Q_t \cdot AC_t(Q_t, W, D) - M(Q_t, W, D) - rK(W, D) \right) \quad (4.1)$$

where $P_t(Q)$ represents the inverse travel demand function expressed in real (price) present value terms; AC_t is the average cost function in year t expressed in present value terms and includes all of the expenses of users including users' value of travel time; the road authority expenses are contained in M , the present value of the average total maintenance cost per year including the cost of resurfacing and K the capital cost of road construction; Q_t the traffic volume in year t ; W the width of the road; D the durability standard (thickness of the pavement) for roads; and r is the real interest rate. The following optimality rules for pricing, investment in capacity and durability standard can be obtained from the following first-order conditions:

$$P_t = \left(AC_t + Q_t \frac{\partial AC_t}{\partial Q_t} \right) + \frac{\partial M}{\partial Q_t} \quad (4.2a)$$

$$T \left(\frac{\partial M}{\partial W} + r \frac{\partial K}{\partial W} \right) = \sum_{t=1}^T Q_t \frac{\partial AC_t}{\partial W} \quad (4.2b)$$

$$T \left(r \frac{\partial K}{\partial D} \right) = \sum_{t=1}^T \left(Q_t \frac{\partial AC_t}{\partial D} + \frac{\partial M}{\partial D} \right) \quad (4.2c)$$

The optimal pricing rule in (4.2a) indicates that the total road charge paid by the user should be equal to the sum of private user costs plus the costs of congestion and road damage repair. This is essentially the social marginal

cost of road use which varies over time. The cost paid for infrastructure would be equal to the sum of the last two terms in 4.2a, $\left(Q_t \frac{\partial AC_t}{\partial Q_t}\right) + \frac{\partial M}{\partial Q_t}$. The optimal road capacity rule in (4.2b) indicates capacity should be constructed to the level where the marginal benefits from reduced congestion delays (the right side of the equation) become equal to the increased capital cost of construction and maintenance (the left side of the equation). Finally, the optimal durability condition in (4.2c) indicates that the durability of infrastructure must be set at the level where the marginal benefit from increasing investment in durability (terms to the right of the equal sign) equals the additional capital cost.

These rules for optimal pricing, capacity and durability can be applied to any modal infrastructure including roads and airports. The optimal pricing and capacity conditions determine prices and investment levels for airports since the cost of increasing the durability of an airport runway is negligible relative to the cost of congestion delays. On the other hand, since road maintenance costs are significantly influenced by the thickness of the road, all three conditions (pricing, investment and durability) need to be determined simultaneously in the case of roads.

4.2 PRICING AND INVESTMENT IN A NETWORK

This subsection examines the issue of transferring funds between parts of a network (aviation, road or rail) in response to the potential sources of so-called “network externalities.” It also discusses arguments that suggest the (allocative efficiency) welfare loss from failing to cross subsidize is minor compared to the (productive efficiency) welfare loss from cross subsidizing.

The essence of the question of pricing and investment rules for multiple roads or airports is whether, for pricing and investment, they should be viewed as independent or as a linked network. If the set of airports, for example, is considered as a network, there may be a price for runway and terminal use which is averaged over the network, similarly to roads. On the other hand, if each airport or roadway is considered independently the price set would consider the cost and demand conditions on that link or at that airport. Many argue that complementarities exist between links (roads) or nodes (airports) and socially efficient pricing must reflect these positive externalities. Averaging prices over the network is one practical way of reflecting such externalities in what is charged. Another method would be to transfer revenue from one facility to another after establishing socially efficient prices.

Consider a system which contains several links and nodes. This could represent a road, air or rail network. Should the setting of user-charges at i be affected by some other node such as j ? For example, if j is a small airport and i is a major airport, can a cross subsidy to j from i be justified on economic grounds? Some would say yes if there is a consumption externality between i and j , or if there are decreasing costs at j and constant returns at i , economic welfare is improved by setting $P_j = MC_j$ and $P_i > MC_i$. It is not clear, however, that either network externalities or scale economies are sufficient to justify a cross subsidy.

Consider a situation in which i and j are operated with optimal productive efficiency (an assumption returned to later) or equivalently they are operating on the most efficient cost function. Set runway prices at i are equal to social marginal cost such that the sum of revenues equals costs. At j , setting price equal to marginal costs results in a deficit. What are the options? If there is some lumpiness at i , the Oum and Zhang (1990) result suggests revenues will be available to subsidize j . A second alternative is to raise fees at i to generate revenues for node j such that the surplus at i equals the deficit at j . In this circumstance the welfare loss would be borne by all routes including movements from j to i . Rather than "tax" users at i , the operator at j could employ Ramsey pricing to achieve full cost recovery at j . Suppose ij traffic is the least price elastic, this group will pay proportionately more than, say, ji passengers. The welfare loss, in this case, would be restricted to the j and i markets. In the former case the welfare loss is spread over i, j, k, n, m and e markets. Which approach has a greater welfare loss will depend on the values of the elasticities and the amount by which price changes. This, albeit, simple analysis leads to the following questions: Is there a rationale for the transfer? Why is there a deficit? Should the set of airports or links be treated as a system rather than as individual operations to be priced separately *and* independently?

First, is there a rationale for the transfer? Some argue that there is complementarity between j and i since j feeds passengers to i . While this is true, the cost savings represent a gain to the airlines providing the service through density economies and are not a gain to the airport. To the extent it is a gain to a private firm, it should be reflected in the landing fees airlines are willing to pay at j . In other words, the density economy gained by the airline would be fully internalized through higher landing fees. It makes little economic sense to transfer the rents to the airline and not have j exploit

the fiscal capacity it has available. This means landing fees should be increased above marginal cost at j in a price-discriminating fashion until the deficit is covered.

Second, why is there a deficit? What about increasing returns to scale or some other form of cost economies which lead to average cost being greater than marginal cost? While this may be a legitimate source of potential gain, it will likely rest with the smaller airport since evidence suggests constant returns to scale for larger airports (Morrison, 1983). Marginal cost pricing results in a deficit at the small airport and breaking even at the large airport. But this is the argument discussed above regarding the relative values of Ramsey pricing at j or increasing all prices at i to cross subsidize j , and there is no evidence one way or the other as to which welfare loss is greater.

Economies of scale at j are not a sufficient reason to cross subsidize from i nor are they a sufficient reason to subsidize from general tax revenues. Indeed, as argued elsewhere, subsidizing from general tax revenues may also result in a net welfare loss. Thus, in the case of airports, there does not seem to be a demand side argument for cross subsidy. The choice is between Ramsey pricing at j or a subsidy from the general taxpayer, whichever has the lower welfare loss. What about roads? The same arguments would seem to hold here as well. Trucks, buses or even cars could have density economies but these should be internalized through normal commercial transactions.

Should the links and nodes be treated as a system or individually? This depends on a number of factors including the second question raised above, why is there a deficit? Other factors include the practical ability to price separately, to have the information to price efficiently, to have some idea of the welfare losses of moving from link- or node-specific prices to some "average" system price and the extent to which there is substitution between links, nodes or routes. The greater the extent of substitution the stronger the case for individualized facility pricing.

Airports have no apparent demand side externalities. They are also sufficiently distinct that it is possible to identify demand and supply at each facility and establish a set of efficient prices. Furthermore, the set of efficient prices leads to optimal investment decisions at each airport to reflect the need for capacity. If there were increasing returns at one or more airports

the decision to “tax” major airports to subsidize smaller airports would have to weigh the welfare gains associated with such a subsidy at the recipient airport against the welfare losses at the taxed facilities. It also seems to imply that there is no competition between airports. To the extent there is competition, the ability to raise revenues for cross subsidy would be reduced.

Moving from a specific to a system fee is probably best illustrated by the system used in Canada until recently; fees were set to be identical for Group I airports regardless of differences in demand and costs between facilities.²⁴ The greater the differences, the larger the welfare loss from system average pricing. Perhaps the greatest argument for individual pricing is to provide the incentives to individual facilities to achieve least-cost production (productive efficiencies) and to exploit available “scope” economies.

There is strong and convincing evidence that system average pricing and a lack of market discipline has led to cost inefficiencies at Canadian airports (Hamilton, 1991). Furthermore, if they are explicitly or implicitly subsidized, airports have little incentive to exploit available opportunities which contribute to commercial success and break-even financial performance. Many small- to medium-size airports in the United States exploit their commercial and concession revenues to a greater extent than similar airports in Canada. They also go beyond producing the products of aircraft movements and enplaned/deplaned passengers; they exploit commercial potential and scope economies. One could argue that the welfare gains available from achieving least-cost operations far outweigh any welfare losses resulting from forcing small airports to be self-financing. Two important characteristics of airports which must be kept in mind are that they are substitutes as well as complements to other airports and they can produce many products besides those conventionally identified with airports.

Similar arguments may be made for roads, but there may be a stronger case for treating roads in a network framework. Like airports, there does not seem to be a convincing case to be made for demand side complementarities which would support an efficiency argument of cross subsidy between links. Different roads have different capacities and thus face different costs and demands. This would favour a pricing structure which reflected these characteristics. It is obvious that the distribution of traffic over the system, and the attendant welfare level, will be quite different if

a single price is charged for the use of any part of the system than if the price is set equal to the marginal cost of each route and node to reflect the cost and demand conditions unique to that part of the system.

The strongest arguments for treating roads in a system are the transactions costs associated with separate pricing of each link and the lack of information regarding demand and costs on each link. It is important, however, that systems be defined to be as homogeneous as possible since, like airports, the greater the divergence in costs and demand across links the greater the welfare loss resulting from system average pricing. Prices will generally be too high on large, high-quality facilities and too low on low-density, low-quality facilities. We may thus end up with too much capacity in the latter because prices greater than marginal cost will attract capital and too little in the former since the low return will discourage capital investment. It is also important to have roadway services, produced in the most efficient way possible, achieve productive economies. This means that some links may need to be abandoned and their freed-up resources devoted to other links yielding a higher return.

4.3 SECOND-BEST PRICING AND SECOND-BEST INVESTMENT

The conventional first-best rule for optimizing the level of investment in transportation infrastructure is the equalization of marginal benefits and costs. If there are institutional constraints which prevent prices from being set at marginal cost in either input or output markets, the optimal investment and pricing rules must be modified. Transportation infrastructure can be financed with tolls (output taxes), fuel taxes (input taxes) or from general revenues. If there are constant returns to scale in production and divisibility in capacity investment, there is no difficulty in efficiently pricing and financing capacity. In the absence of these two conditions, however, and given the constraint that users should pay the full cost and finance the facility without a subsidy from general revenues, optimality requires a second-best investment rule in conjunction with second-best prices (input and output taxes). As Friedlaender and Mathur (1982) have shown, these investment levels and rules depend on the nature of the price distortions and financing constraints.

A second-best investment rule differs from the first-best rule. With first-best pricing and investment, prices are set equal to marginal cost. There are no output or input taxes, and investment is carried to the point at which the

marginal cost savings resulting from the investment in infrastructure equals the marginal cost of the infrastructure investment. In a second-best situation the investment rule depends on the way in which the infrastructure is priced and whether there is a financing constraint. Two quite different cases can occur.

In one state, the problem may not be one of financing the facility, insuring total revenue equals total cost, but rather one in which there are output market distortions such as congestion or vehicular-generated pollution. This means that in the absence of an externality charge the mix of prices will not yield an economically efficient outcome. If there is underpricing, for example, the price set for roadway use is less than marginal social cost. A second-best pricing and investment strategy would set a fuel tax and reduce the size of the infrastructure relative to a first-best level. On the other hand if an "arbitrary" fuel tax is in place, as when governments use fuel as a source of general revenues, the second-best rule leads to setting infrastructure prices at less than marginal social cost yielding an implicit subsidy, as well as again reducing the investment in infrastructure capacity.

In this second-best situation, the reduction in the supply of capacity is used to reduce the efficiency loss resulting from the distortion in prices. This is true as long as fuel and infrastructure are substitutes (Friedlaender and Spady, 1981). If fuel prices are too high, people substitute towards infrastructure since its relative price has fallen. To offset this distortion, the supply of infrastructure is reduced to make it relatively more scarce and increase its price. Furthermore, to the extent that the elasticity of facility use with respect to the size of the facility is greater than zero, a reduction in investment brings about a reduction in demand. The same arguments hold when there is underpricing.

A second case, and one perhaps more relevant to the discussion at hand, occurs when there is a financing constraint. The long-run pricing and investment problem is to define an optimal structure of infrastructure prices and fuel taxes and determine the optimal investment in capacity. The difference in the level of investment between the first- and second-best investment rule depends on the sensitivity of revenue with respect to the amount of capacity and output taxes. Revenue will be sensitive to the level of capacity because, as capacity increases, the number of users increases and revenue from output prices will rise. If usage is highly elastic with respect to the level of capacity revenue will also be elastic with respect to infrastructure.

Friedlaender and Mathur (1982) showed that if revenue and usage are not particularly elastic with respect to investment, it is desirable to reduce investment relative to the first-best level because distortions caused by second- best user-charges, needed to finance the capacity, generate a deadweight loss burden. If revenue is elastic with respect to capacity, as it may be at some hub airports, it is desirable to expand capacity beyond the first-best level.

The second-best investment rule provides a linkage between:

- the optimal difference between the marginal benefits and costs of the infrastructure and the sensitivity of revenues to output prices and fuel taxes; and
- changes to the level of infrastructure capacity.

The simulations which have been undertaken to examine the welfare differences of first- and second-best investment rules in the presence of pricing distortions or financing constraints have consistently illustrated that the welfare loss of using first-best investment rules is relatively small. Friedlaender and Mathur (1982) undertook the simulation for rail and road and Borins (1978) performed a similar analysis for air. In both cases, the welfare surface is relatively flat primarily because the benefit function is relatively flat. Therefore, second-best pricing rules in practice can rely on first-best investment rules in establishing the optimal investment in capacity.

4.4 USER-CHARGES AND COST RECOVERY

Moving from the principles, outlined in Section 2 of this paper and formalized in subsection 4.1, to practice in pricing transportation infrastructure requires the development of pragmatic schemes which approximate or build upon social marginal cost pricing while simultaneously providing predictability to users and capital markets. Practical pricing schemes are discussed below, and some information of the type of changes which the various user-groups in the modes might expect with a move from current pricing principles to more economically efficient ones is provided. The illustrations are not meant to be taken as prices which would be implemented tomorrow if there is agreement to change the pricing principles. Rather they are designed to be illustrative of the magnitude and type of changes for different facilities and different user-groups with a shift to more efficient pricing.²⁵

4.4.1 Airport User-Charges

Runways at large airports are generally built to handle the largest and heaviest aircraft currently in use by commercial air carriers. The additional construction costs for increasing the durability of a runway is small relative to the congestion delay costs.²⁶ Therefore, the condition of optimal durability in equation (4.2c) can be ignored in discussing airport pricing and investment.

Major Airports

The 1991 landing fee schedule for most Canadian airports is reported in Table 4.1. The same fee schedule is applied to all Group I and II airports and major international airports operated by Transport Canada at all times, including peak periods. The only exception is that minimum landing fees were imposed starting in 1991 at Vancouver and Toronto international airports. The use of uniform landing fees at all airports is both inequitable and inefficient because both social marginal cost and price elasticities differ significantly across Canadian airports. Other than charging higher landing fees to international flights, the current landing fees do not distinguish between short and long flights.²⁷

Table 4.1
LANDING CHARGES
GROUP I, II AND MAJOR INTERNATIONAL AIRPORTS
CANADA

Domestic Flights		
Charge: \$/1000 kg or fraction thereof		
Weight (kg)	Jet	Turboprop
Less than 21,000	\$2.27–2.52	\$1.92–2.13
21,000–45,000	\$2.87–3.20	\$2.33–2.60
More than 45,000	\$3.40–3.78	\$2.87–3.20
International Flights		
Charge: \$/1000 kg or fraction thereof		
Weight (kg)	Jet	Turboprop
Less than 30,000	\$2.94–3.27	\$2.39–2.66
30,000–70,000	\$3.63–4.03	\$3.15–3.51
More than 70,000	\$4.01–4.58	\$4.38–4.87

Source: Transport Canada (1991).

For major airports such as Toronto, Vancouver, Calgary and Dorval Airport in Montreal, external congestion costs are a major component of the social marginal costs for runway use. Therefore, congestion tolls could become the major source to finance capacity expansion. However, the level of congestion varies substantially by the time of day. This means that peak-period pricing may be an attractive means to approximate social marginal cost pricing for runways and terminals. For example, the British Airports Authority (now privatized as BAA Plc.) has implemented peak-period pricing at all of its seven airports. At major airports such as London's Heathrow, the peak-period landing and terminal fees are about five or six times higher than the off-peak fees. The implementation of peak-period pricing in Canada could lead to a dramatic change in the structure of landing/takeoff and passenger terminal fees at major Canadian airports where congestion problems exist. Peak-period pricing would result in a dramatic increase in peak-period landing and passenger terminal fees, the elimination of the current weight-based landing fees, and the reduction of off-peak landing fees below their current levels.

Table 4.2 illustrates the primary results of a simulation of introducing peak-period pricing at Pearson International Airport (PIA), (Gillen, Oum and Tretheway, 1988). The authors examined the hourly and the daily traffic variations at PIA and found that a three-part pricing schedule made sense; one for weekday evening peak hours (1700 to 2100), one for weekday mornings (0700 to 1000) and weekend evening peak hours (1700 to 2100), and one for all other times, also referred to as off-peak hours. The charges are highest during high peaks, somewhat lower fees are levied in the lower peak, and the smallest fees are assessed in the off-peak periods. Further, an analysis revealed two distinct seasons that dominate aircraft movements: March–October, referred to as the high season and November–February, referred to as the low season. They, therefore, recommended an annual fee structure of six prices, one for each of high-peak, low-peak and off-peak periods for the two (high and low) seasons.

The authors estimated social marginal costs and compared them with the landing fees for selected aircraft in 1985 (See Table 4.2). However, under social marginal cost pricing, general aviation (GA) aircraft (represented in their analysis by light piston aircraft) which previously paid nothing for the use of runways,²⁸ would pay \$31 for landing during off-peak periods, \$161 during the summer high-peak, and \$43 during the winter low-peak periods.

A heavy aircraft such as a B747-200, which used to pay \$769 and \$521 for a takeoff/landing of international and domestic flights, respectively, would pay \$426 during summer high-peak, \$246 during summer low-peak, \$271 during winter high-peak, \$226 during the winter low-peak and \$206 during off-peak periods. The implication is that social marginal cost pricing would reduce the landing and takeoff fees for large aircraft, while substantially raising the fees for small aircraft, especially during peak periods, from their current values.

Table 4.2
COMPARISONS BETWEEN SOCIAL MARGINAL COSTS AND 1985 LANDING FEES FOR SELECTED AIRCRAFT, PEARSON INTERNATIONAL AIRPORT, TORONTO (1988 CAN \$)

	Aircraft type					
	B747-200	DC10-40	B737-200	Dash 8	Business jet	Piston
1985 fees						
Domestic	\$521	\$355	\$80	\$12	N/A	\$ 0
International	\$769	\$524	\$85	\$15	N/A	\$ 0
Social marginal cost						
<i>High months</i>						
High-peak	\$426	\$376	\$269	\$213	\$211	\$161
Low-peak	\$246	\$196	\$154	\$ 98	\$105	\$ 55
<i>Low months</i>						
High-peak	\$271	\$221	\$170	\$114	\$120	\$ 70
Low-peak	\$226	\$176	\$142	\$ 86	\$ 93	\$ 43
Off-peak	\$206	\$156	\$129	\$ 73	\$ 81	\$ 31
Price elasticity	N/A	-0.068	-0.075	-0.086	N/A	-0.58

Source: Gillen, Oum and Tretheway (1988).

The implementation of peak-period pricing based on social marginal costs by type of user is likely to shift demand for airport services to less congested airports. Piston engine aircraft which represent the bulk of the general aviation traffic could be virtually eliminated during high and low peaks in terms of the number of landings. Furthermore, Gillen, Oum and Tretheway’s estimates of the price elasticities indicated that peak-period pricing would probably bring some marginal changes to air carrier scheduling. Commuter

carriers would make the most changes in their scheduling to avoid the high-peak period fees; thus, the movement of heavier aircraft would increase slightly as the charges decrease.

If social marginal cost pricing had been implemented in 1986, PIA's revenue from runway operations could have doubled to reach \$26.6 million. This would allow PIA's airfield operation to generate surpluses, after covering the interest and depreciation expenses. This surplus could be used to expand capacity when economically efficient to do so.

Small Airports

Fees at small airports can also be based on social marginal cost. However, since there is very little congestion at small airports, peak-period user-charges would not be much different from off-peak periods. Therefore, (social) marginal cost pricing would most likely lead to financial deficits since marginal cost would be less than average costs for the reasons discussed in Section 2. The two alternative modifications to efficient prices which allow substitution of a cost-recovery target were discussed in Section 2. First, the Ramsey pricing principle may be applied to achieve a cost-recovery target (including break even). Since the size of the quasi-optimal markup over marginal cost is determined on the basis of willingness to pay (that is, price elasticity), it makes sense to charge higher landing/takeoff fees to larger aircraft, which carry higher payloads than small aircraft. Perhaps this was the economic rationale for the weight-based landing fee structure currently in use in many countries, including Canada. The weight-based fee structure loses its economic rationale when congestion develops. Second, a two-part tariff (usage/access charges) can be applied to cover the system-wide or individual airport shortfall in revenue by charging access (licence) fees while charging site-specific social marginal costs as usage fees.²⁹

In principle, the choice between Ramsey pricing and two-part pricing structures depends on the characteristics of demand and the proportions of fixed and variable cost in total cost. It is these elements which effect the welfare gains. Ramsey pricing would be preferred over two-part pricing structures if the elasticity of entry to the system or access to an airport is non-zero. If access is price sensitive, the variable portion of the multi-part price must be increased above marginal cost. In the case of multiple users, this increase would rely, essentially, on a Ramsey concept. On the other hand, usage/access

fees would be preferred, from a welfare gain perspective, if the elasticity of demand with respect to access were zero, or equivalently, demand were fixed.

Theoretical second-best and “applied” second-best pricing may lead to the selection of different pricing principles, at least in the short run. An evolutionary rather than revolutionary approach may be required to minimize the complexity and to gain experience with a new system. Both users and suppliers have time to adjust. If a new pricing principle creates uncertainty, it may, in the end, result in a net welfare loss. There is a balance between moving so quickly that markets are disrupted and the potential welfare gains are lost, and moving so slowly that incumbents become entrenched, and institutions are unable to fully implement socially efficient pricing, again resulting in foregone welfare gains.

London airport, a small national airport in southwestern Ontario, is an example of the consequences of introducing new pricing schemes at small airports. The airport depends largely on the presence of the military. It is the home base of a commuter/feeder carrier. In the 1984-85 fiscal year, the airport recovered only about 20 percent of the total airfield costs from landing fees. Gillen, Oum and Tretheway (1986) illustrated the application of both social marginal cost pricing and the Ramsey price principle by type of user for this airport. Table 4.3 is taken from their study. The upper portion of the table summarizes 1985 revenues from the landing fees.

The mid-section of the table shows that application of marginal cost pricing by type of user would eliminate piston aircraft and severely reduce turboprop movements. Jet aircraft movements would also be reduced but only slightly. The marginal airfield costs estimated for this airport are \$193, \$110 and \$78 for jet, turboprop and piston aircraft, respectively. These compare to the current average landing fees of \$164, \$20 and \$0. Under social marginal cost pricing, the airfield revenue was expected to reach approximately \$862,000 which is higher than the actual 1985 revenue. Nevertheless, it only represented about 39 percent of the 1985 airfield costs.

To achieve higher cost recovery with minimum loss of economic efficiency, a Ramsey pricing principle can be implemented. Ramsey pricing with a 62.5 percent cost-recovery target, for example, requires charging \$361, \$166 and \$85 to jet, turboprop and piston aircraft, respectively. The results also

show that implementing these charges would nearly eliminate piston aircraft movements and reduce jet and turboprop landings to 1,788 and 4,464 from the actual landings of 1,856 and 6,512, respectively. Total revenue could reach \$1.4 million, about 62.5 percent of the 1985 actual airfield costs.

Table 4.3
SOCIAL MARGINAL COST PRICING AND RAMSEY PRICING BY TYPE OF USER FOR LONDON (ONTARIO) AIRPORT

	Jet	Turboprop	Piston	Total
1. 1985 situation				
No. of landings	1,856 (7.3%)	6,512 (25.7%)	16,982 (67.0%)	23,350
Average landing fee estimates	\$163.66	\$ 20.0	\$ 0	—
Revenues	\$303,760	\$130,240	\$ 0	\$434,000
1985 airfield cost	—	—	—	\$2,223,000
2. Marginal cost pricing				
Estimated MC	\$192.51	\$110.29	\$ 77.94	—
Price elasticity	-0.04	-0.065	-0.325	—
No. of landings (forecast)	1,843	4,601	0	6,444
Revenue (forecast)	\$354,778	\$507,398	0	\$862,176
Cost recovery ratio	—	—	—	39%
3. Ramsey pricing with estimated marginal costs				
Ramsey prices	\$360.95	\$166.41	\$ 85.29	—
% Markup	87%	50%	9.4%	—
No. of landings (forecast)	1,788	4,464	0	6,252
Revenue (forecast)	\$645,378	\$742,854	0	\$1,388,232
Cost recovery	—	—	—	62.5%

Source: Gillen, Oum and Tretheway (1986).

Again it should be noted that costs are expected to fall as airports are defederalized. Finally, it is important to point out that marginal social cost pricing may not result in deficits at small airports if there is constant returns to scale and no significant lumpiness in capacity investment.

4.4.2 Summary of Pricing and Airport Cost Recovery

The major conclusions reached and changes advocated with respect to the current user-charge system may be summarized as follows:³⁰

- The uniform user-charge system should be repealed in favour of site-specific user-charges.
- The current weight-based landing fees result in fees which are too low for small aircraft (primarily general aviation and corporate aircraft) and too high for large aircraft during congested periods.
- Peak-period user-charges should be implemented in airports with a congestion problem, and the differential user-charges between peak, low-peak, and off-peak periods should reflect their respective social marginal costs, including congestion externality costs. This tends to lead to a large differential between peak and off-peak fees.³¹
- A weight-based landing fee, if appropriately set, may be consistent with the Ramsey pricing principle, and thus may be economically justified at small uncongested airports. This is because the demand for airport services by larger (heavier) aircraft is less price-elastic than by smaller (lighter) aircraft. The rationale for weight-based pricing breaks down, however, when the airport becomes congested and congestion pricing is applied.
- Charging higher landing fees for long-haul international flights at uncongested airports can be economically justified because it is consistent with the spirit of Ramsey pricing; the demand for airport services by long-haul flights is more price-inelastic than short-haul flights.

4.5 ROAD USER-CHARGES

Previous research has led to the conclusion that the best way to economize on maintaining and using an existing road is to set user-charges equal to social marginal costs, that is, the actual cost each user imposes on society, including the effect on the road's condition, noise, pollution and the delay imposed on others. Such a charge would ensure that independent decisions by users reflect the interests of society. On the other hand, the social cost of road use largely depends on the design of the road. Road infrastructure provides capacity in the form of traffic lanes and durability in the form of pavement thickness to facilitate its use by heavy vehicles. Both road capacity

and durability investments are expensive and involve scarce resources. The scarcity associated with capacity causes congestion costs while the scarcity associated with durability causes greater costs for road wear and damage (road deterioration) than for the construction of more durable roads.

Road charges should differ by type of user; for example, heavy versus light vehicles, passenger versus freight services or commercial (common carriers) versus private users. The empirical results of two previous studies on this topic are summarized to provide an indication of the direction of adjustment required in road user-charges, particularly by type of road users: a study by Nix (1989) entitled *Road-User Costs* using the most recent Canadian data, and a U.S. study undertaken by Small, Winston and Evans (1989) entitled *Road Work*.

4.5.1 Nix (1989)

The most significant attempt to estimate road-user costs in Canada was by Haritos (1973). He used 1968 data and employed engineering and regression approaches to allocate road costs. Nix (1989) updated Haritos' study with more recent data compiled by the Road Transport Association of Canada (RTAC) and up-to-date information on vehicle characteristics.

Haritos adopted a two-part price structure: maintenance costs (regarded as annual costs) and capital costs (fixed in the short term). The costs considered avoidable within the time frame of one year were considered escapable and, therefore, included in the (short-run) marginal cost. This measure should be used to set the user-charge per vehicle-kilometre. The costs not avoided within a year are considered inescapable, and thus treated as a fixed cost in the short run. Haritos argued that short-run fixed costs should be recovered through annual fees such as vehicle registration and licence fees.

Table 4.4 summarizes the results of Nix's cost allocation study.³² It compares the 1986 road user-charges and the road costs by vehicle category. The results show that the annual fixed user fees (licence fees) for cars and trucks are too low to cover the short-term fixed cost of roads. In fact, the revenues from the annual fixed user-charges cover only about one third of the fixed road costs. A comparison of the user-charges and road costs imposed by user group (Table 4.4) indicates that charges for passenger cars and other light vehicles are much higher than the financial cost they impose

on the road by a factor of 3 to 1 while charges are too low for trucks. In other words, light vehicles, which do not cause much damage to roads, are charged three times the cost per kilometre while heavy vehicles are undercharged. This result may not be interpreted as overcharging light vehicles because neither Nix nor Haritos considered the externality costs of noise, air pollution and congestion delay a driver imposes on others. However, it is safe to interpret the result as undercharging heavy vehicles. Nix also pointed out that, during the past two decades, road user-charges in Canada did not increase as fast as road construction costs.

Table 4.4

COMPARISON OF ROAD-USER-CHARGES AND COSTS^a
(1986 CAN \$)

	User-Charges		User-Costs	
	Range	Mid-point	Range	Mid-point
Annual charges versus annual costs				
Cars	\$86		\$253-\$316	\$284.5
3-axle truck (18.1 t)	\$250-\$675	\$462.5	\$1881-\$2606	\$2243.5
3-S2 tractor (32.1 t)	\$450-\$1600	\$1025	\$2897-\$4014	\$3455.5
3-S3 tractor (46-49 t)	\$600-\$2700	\$1650	\$4140-\$5739	\$4939.5
8-axle B-train (62.5 t)	\$1450-\$4000	\$2725	\$5394-\$7478	\$6436
B-train, empty (18.1 t)	\$1450-\$4000	\$2725	\$1575-\$2182	\$1878.5
Trip charges (fuel) versus trip costs, per kilometre^b				
Cars	\$0.012-\$0.020	\$0.016	\$0.004-\$0.006	\$0.005
3-axle truck	\$0.051-\$0.084	\$0.068	\$0.026-\$0.056	\$0.041
3-S2 tractor	\$0.069-\$0.097	\$0.083	\$0.043-\$0.090	\$0.067
3-S3 tractor	\$0.071-\$0.117	\$0.094	\$0.063-\$0.130	\$0.097
8-axle B-train	\$0.086-\$0.141	\$0.114	\$0.084-\$0.172	\$0.128
B-train, empty	\$0.051-\$0.084	\$0.068	\$0.020-\$0.046	\$0.033

Source: Nix (1989).

- a Annual costs include the appropriate annual vehicle charge plus an annual charge for each axle on the vehicle.
- b Trip cost includes the appropriate vehicle cost per kilometre given the vehicle weight plus a cost per axle-kilometre for each axle.

The Nix study represents an important conceptual contribution. Clearly, there may be some argument regarding the exact numbers since those reported are based on a particular set of assumptions. What the paper does indicate, however, is that moving to marginal social cost pricing of roads has a relatively greater impact on the structure of prices among road users than on the level of prices. Indeed, with the recent increases in fuel taxes by provincial governments, there is a real possibility that car users are paying their full social costs of road use on uncongested roads. Changing the basis of prices to social marginal costs will affect the distribution of costs between car and truck and most significantly among different types of truck. The following study reinforces this point.

4.5.2 Small, Winston and Evans (1989)

Small, Winston and Evans derived road user-charges under an optimal investment policy (with respect to road capacity and thickness of pavement), taking into account the combined cost to the vehicle owners and infrastructure providers. Inclusion of the cost to vehicle owners of differences in highway capacity and durability was the distinguishing feature of this study as compared to Haritos (1973) and Nix (1989). This allowed Small, Winston and Evans to take into account the effects of a lower standard of road maintenance on the cost of vehicle maintenance and fuel consumption.

The two primary considerations in pricing the use of existing roads are road wear costs and congestion delay costs. The road wear costs include road maintenance costs and the user-costs for operating vehicles, including vehicle repair, vehicle depreciation, fuel expenses and their value of time. The authors computed optimal road durability by minimizing the sum of annualized road maintenance costs, user (passengers and carriers) costs and capital costs for road construction. This procedure is equivalent to determining the optimal level of investment in road quality (thickness) through a series of cost-benefit analyses. The study found that the United States builds roads of lower quality (too little thickness) than economically optimal. This increases both the combined road infrastructure and operator costs, making the road transportation sector less cost efficient than would be the case if roads were built to optimal standards. Computations are made for rigid as well as flexible pavements.

An important conclusion of the study was that a United States road-pricing policy should be targetted to reduce weight per axle as vehicles with heavy loads per axle cause most of the road damage. They emphasized that it is the weight per axle that matters, not total vehicle weight. For example, a 50,000-pound, two-axle truck causes more road wear than a huge twin-trailer rig spreading 100,000 pounds over seven axles. Most road damage is caused by heavy vehicles with a small number of axles.

To measure the congestion costs caused by vehicle traffic, the study used the concept of passenger car equivalents (PCE) per hour. The PCE for each vehicle was determined by the amount of road space it effectively took up, including the space between vehicles required for safety, compared with that of an average car. For example, a typical truck or bus has two-to-five passenger car equivalents. As in Nix's study, the authors relied on an engineering approach to quantify the relationship between traffic volume and speed. This relationship is essential to measure the extra delay caused by adding one passenger car equivalent to the traffic stream. The authors advocated significant changes in road-pricing policy for heavy trucks, shifting from a reliance on fuel taxes and weight-graduated licence fees to one of direct mileage charges steeply graduated with respect to axle loads. They estimated that such a pricing system in conjunction with a modest increase in capital outlays on improving road thickness ("a sensible investment policy") could reduce maintenance costs by about \$9.4 billion a year.

Table 4.5 presents the estimates of the effects of their policy on the maintenance costs and "tax revenue" (user-charge revenue) attributable to each truck type, as a share of the total maintenance costs. This table shows that for most truck categories current user-charge revenues fall far short of maintenance costs attributable to them. The user-charge ("tax") revenues for intercity and urban roads cover only 29 percent and 14 percent of pavement maintenance costs, respectively. In contrast, under the proposed policy, the total maintenance costs for both intercity and urban roads would be fully recovered by user-charge revenues. Furthermore, the authors showed that the proposed policy would result in a welfare gain of \$8 billion a year over the current pricing and investment policy.

Table 4.5

CONTRIBUTION TO ALLOCABLE MAINTENANCE COSTS BY USER-TYPE^a

Vehicle type	Current pricing and investment		Optimal pricing and investment	
	Share of allocable maintenance costs (%)	Tax revenue contribution to allocable maintenance costs (%)	Share of allocable maintenance costs (%)	Tax revenue contribution to allocatable maintenance costs (%)
Intercity				
SU2	45.19	9.81	10.07	19.38
SU3	0.74	1.31	1.33	1.35
TT4	0.43	0.11	0.21	0.21
TT5	0.19	0.09	0.33	0.33
CS3	2.11	0.63	1.70	1.71
CS4	0.76	0.33	0.72	0.72
CS5	43.02	13.51	60.32	60.38
CS6	6.34	2.83	14.60	14.62
DS5	1.18	0.42	1.68	1.68
DS6	0.04	0.01	0.04	0.05
Total	100.0	29.05	100.0	100.43
Urban				
SU2	88.24	11.88	67.52	67.92
SU3	3.10	1.25	9.61	9.67
TT4	0.61	0.13	0.96	0.96
TT5	0.06	0.01	0.09	0.09
CS3	0.58	0.18	2.49	2.50
CS4	0.25	0.07	0.61	0.61
CS5	5.82	0.40	6.95	6.96
CS6	0.90	0.13	10.51	10.52
DS5	0.38	0.07	0.87	0.87
DS6	0.06	0.01	0.39	0.39
Total	100.0	14.13	100.0	100.49

Source: Small, Winston, and Evans (1989).

- a Type of vehicles as follows: SU2, Single Unit 2-Axle; SU3, Single Unit 3-axle; TT4, Truck Trailer 4-axle; TT5, Truck Trailer 5-axle; CS3, Conventional Semi 3-axles; CS4, Conventional Semi 4-axle; CS5, Conventional Semi 5-axle; CS6, Conventional Semi 6-axle; DS5, Double 5-axle; DS6, Double 6-axle.

4.5.3 Summary of Road Cost Recovery by User-Group

The studies conducted in Canada and the United States indicate:

- Passenger cars and other light vehicles pay a disproportionately higher share of the total road maintenance costs as compared to trucks and other heavy vehicles.
- Lack of congestion tolls as well as other externality charges in both countries indicate that the current user-charge system undercharges road users on congested roads. It may also undercharge with respect to environmental costs. This is more likely in the United States where fuel tax rates are less than half of those in Canada. With increases in fuel taxes in the last year, it may well be that car users, in Canada, are paying approximately their marginal social costs on uncongested roads.
- Both the Canadian and American studies illustrate that an economically efficient pricing scheme would affect both the average level of road prices *and* the burden across user-groups. If one interprets a fair or equitable pricing scheme as one in which the prices users are paying reflect the costs which they impose, then the efficient pricing scheme which shifts the burden can also be said to be fair.
- Total road costs (the sum of capital and maintenance costs) are sensitive to the level of investment in capacity and durability and to environmental factors. In the United States, the underinvestment in durability (that is, building sub-standard roads) has resulted in an economic loss amounting to billions of dollars annually due to the increased road maintenance costs and damage to the vehicles. The study by Nix, Boucher and Hutchinson (1992) indicated that roads in Canada are built to a higher level of durability due to the harsher climate. Road deterioration costs range between 50 percent and 80 percent of total costs. Therefore, the Small, Winston and Evans (1989) numerical results must be interpreted with some care when placed in a Canadian context.

5. ALTERNATIVE METHODS OF FINANCING ROAD INFRASTRUCTURE

5.1 INTRODUCTION

In this section alternative methods of financing road infrastructure proposed by analysts and practitioners are described and evaluated. The evaluation is

based on several criteria: the effect on the efficient use of existing capacity, the effect on the efficiency of investment, the implications for equity, administrative feasibility, ease and cost of collection, the practicality of generating revenue and domain of applications (that is, national, local or project basis).

For discussion, the charges related to the road transportation sector are classified into three major groups: vehicle usage, vehicle acquisition and ownership and beneficiaries of road access.³³ The most widely applied user-charges are the motive fuel taxes and annual registration and licence fees. These, along with the general revenue of the governments are the traditional sources of road financing in most countries.³⁴ This subsection describes each of these categories of payments made by road users. An evaluation of each category as a road financing method is treated in subsection 5.2. And finally, several alternative sets of the charging instruments are discussed and evaluated in subsection 5.3.

5.2 THREE CATEGORIES OF CURRENT USER-PAYMENTS

Charges Related to Vehicle Usage

Fuel taxes form a major portion of the total revenue collected from the road transportation sector. In 1989, they accounted for 77 percent of the total revenue from the road sector in Canada. The use of fuel taxes can be justified in a number of ways including road usage, energy conservation and environmental pollution. The level of fuel tax can have an effect on vehicle usage, as well as the number and type of vehicles owned. Fuel taxes provide a cost-effective and administratively feasible means of allocating variable charges. There are some who argue that they approximate road usage costs for uncongested intercity roads because the total amount of tax paid varies in direct proportion with road usage. Whether fuel taxes are a reasonable proxy for reflecting road costs depends on the type of roadway and the level of traffic.

Charges Related to Vehicle Acquisition and Ownership

This category includes vehicle licence fees, registration fees, vehicle inspection fees, vehicle transfer taxes and excise taxes on the purchase of vehicles, etc. Licence fees can be substantial and are usually used in conjunction with fuel taxes. In some countries, fairly steep taxes are imposed when a car is acquired. The revenues from annual licence fees and payments related to vehicle acquisition usually contribute to those road costs which do not vary,

in the short run, with road use. Other payments in this category are generally regarded as fees for the services rendered by government, and thus, are not considered as a source of financing for roads.

Charging Beneficiaries of the Improved Road Access

The construction costs associated with access roads, in Canada and the United States, have usually been financed from general tax revenues (of both the central and the local governments). Local improvement taxes have been used in the past. Now, as their current fiscal position deteriorates, governments are attempting more and more to charge these construction costs against beneficiaries. This approach takes various forms. In the case of new subdivisions, some governments require developers to build the roads and to recover their costs from subsequent purchasers. For major improvements to access roads into a city, some local governments levy land-value increment taxes in the district clearly benefiting from the project (special assessment district). In their recent study, Allen and Floyd (1991) concluded that the tax on special assessment districts along with toll roads were the only promising new sources for funding large-scale expansion projects.

5.3 EVALUATION OF THE ALTERNATIVE FINANCING TOOLS

The pros and cons of each of the charging instruments are evaluated below: motive fuel taxes, a graduated per-kilometre tax based on axle weights, congestion tolls, construction of toll roads, licence fees and taxes associated with acquiring and owning a vehicle of given type, and the methods of charging beneficiaries of the improved road access.

Motive Fuel Taxes

Administratively, fuel taxes are easy to collect and can be used to generate large sums of money for road infrastructure as fuel demand is relatively price-inelastic. At first glance, fuel taxes also appear to be economically efficient because the more extensively road infrastructure is used, the higher the total amount of fuel tax the user ends up paying. Fuel taxes provide a relatively efficient and practical means of pricing road use but only for cars and only for uncongested facilities. Although a high fuel tax (the case in many western European countries) may have some effect in discouraging low-value road users, there are at least two important reasons why fuel taxes are not as efficient as they appear for congested urban and near-urban

roads. First, the largest component of the social cost on these facilities is the cost of the congestion externality. Fuel consumption, however, is essentially unrelated to the extent of congestion externality.

Second, as Small, Winston and Evans (1989) pointed out, usage of motive fuel taxes as a major source of road financing encourages operators to use heavy trucks with fewer axles (that is heavier loads per axle). When the load per axle increases, the amount of damage to the pavement also increases exponentially (Winston, 1991). The amount of fuel tax a trucker pays under the current system rises with a vehicle's axles since trucks with more axles require larger engines and obtain lower fuel economy. This provides truckers with an incentive to reduce the number of axles thereby increasing the load per axle. These arguments make it clear that the current fuel tax system is not an economically efficient way to finance road systems. The fuel tax is neither efficient for pricing road damages (maintenance cost), nor is it efficient for pricing congestion externality. Therefore, the mechanism for pricing road use must differ across user-types. To restrict usage fees for all users to the fuel tax may well result in a net welfare loss.

Graduated Per-Kilometre Tax Based on Axle Weights

As a way to price road damages efficiently, Small (1990) argued for revamping the current U.S. system of road taxes, and adopting a "steeply graduated per-mile tax based on axle weights."³⁵ This pricing system for road damages makes economic sense because as load per axle increases the amount of damage to the pavement increases exponentially. Furthermore, an extensive study by the U.S. Federal Highway Administration found such a graduated tax system to be feasible administratively. It requires slightly more record keeping than the weight-distance charges being used in Iceland, Norway, Sweden and several states in the United States. Small points out that such a graduated per-mile tax, which takes into account both weight and axle configuration, has been in effect, on a systematic basis, in New Zealand for years.

Several recent studies including Nix, Boucher and Hutchinson (1992) and Nix (1989) have shown that a substantial portion of the road maintenance and resurfacing costs in Canada is attributable to "environmental factors" such as elapsed time, weather and other climatic conditions. Therefore, assuming no interaction between deterioration due to environmental factors or due to use, implementation of the steeply graduated per-kilometre tax

based on axle weight alone is not going to achieve maximum efficiency. For maximum efficiency, the portion of road maintenance and resurfacing costs attributable to the environmental factors may be recovered by a lump-sum tax such as annual vehicle licence fees, which may not influence usage of roads significantly depending on the value of particular elasticities.

How the fixed or unallocatable costs are allocated between a variable charge and a fixed charge depends on the relative values of three elasticities: the elasticity of access with respect to the fixed fee, the elasticity of access with respect to the variable fee and the elasticity of usage with respect to the fixed fee. If usage and access are completely inelastic with respect to the lump-sum or access fee, this fee should be set equal to average fixed costs. If, however, the just mentioned elasticities are non-zero, economic efficiency dictates that the fixed fee be reduced and the variable fee be increased. Indeed, one may expect that a greater part of the burden of meeting the revenue requirements is placed on the variable charge rather than the fixed charge since it provides more opportunity for those with more elastic demands to avoid the higher charge by consuming less of the good yet still have the opportunity to consume some of the good. A high access fee may preclude a number of people from the market. Clearly, a balance is required since too high a usage charge may also shift users to other markets.

Congestion Tolls

For nearly a century, transport economists have advocated that road authorities implement congestion tolls. The basic analytical framework for peak-period pricing (based on congestion tolls) was pioneered by Pigou (1912) and modelled formally in a short-run framework by Walters (1961). As discussed, Mohring and Harwitz (1962) recast the analysis into a long-run framework and established the relationship between optimal tolls and cost recovery. Other economists who have worked on the subject have recommended the use of peak-period pricing, as a practical means to charge for congestion or variations in demand with fixed capacity. Peak-period pricing has traditionally been advocated in those cases where there is no interaction between the users of a facility. Thus, the use of the facility by one user does not impose a cost on other users such as with congestion. For example, residential users demanding electrical power during the evening do not impose a cost on industrial users who demand it during the day. Similarly, car users of a highway in the evening do not interfere with car users during the day, because in both cases the demands are independent.

Congestion tolls are recommended when the use of a facility by one user imposes a cost on other users. For example, the use of a runway or approach path by one aircraft means that other aircraft cannot use it and must incur a cost by waiting. The demands are interdependent. In the case of congestion there is a difference between private cost and social cost whereas, with peak loads, private and social costs are the same but the issue is one of who should carry the burden of capacity costs. When peak-period pricing is used as a practical alternative, it is being used to proxy the difference between private and social costs.

The chief advantage of using congestion tolls to finance the road system lies with the improved economic efficiency it achieves as it results in users paying the social marginal costs of their transportation choices. This would not only lead to the efficient use of the given capacity in the short run but also act to provide a signal for efficient capacity investment programs in the long run. Despite these advantages, except for the electronic road pricing experiment in Hong Kong (see Hau, 1990),³⁶ no government uses congestion tolls as a major source of financing for road infrastructure.

In the past, several problems were cited as barriers to the popular acceptance of congestion tolls. First, it is not a straightforward case of calculating socially optimal congestion tolls by road sections and by time of day or day of week. Since congestion consumes enormous amounts of a valuable resource (people's time), a reasonable approximation, which can be done quite easily, is likely to be better than completely dismissing this sound pricing method. The second barrier is that it used to take time (at least to slow down) to collect tolls. This problem is now largely solved since electronic vehicle identification (EVI) technology has become quite reliable, as proven through the Hong Kong experiment. Heggie (1991) indicates that the advent of electronic toll collection can reduce the costs of toll collection on inter-urban toll roads and bridges to under 5 percent of gross revenues. Given the rapidly advancing technology in electronic vehicle identification, toll billing and collection, congestion tolls are likely to become an important new source for road financing.

The third barrier is the alleged unfair distribution of benefits. Opponents of congestion tolls indicate that poor working people and downtown business interests have the most to lose under a congestion charging system. Careful use of the congestion toll revenues may compensate those who are

expected to lose under the system. For example, Small (1990) argued that, in the case of major urban areas such as San Francisco, Los Angeles or Toronto, revenues from congestion tolls would be so large that there would be little doubt that it is possible to fully offset the effects of peak user-charges on nearly all groups, including poor working people and downtown business interests. It is also possible to eliminate the need for fuel taxes, registration fees and a large portion of the local property tax and sales tax funds now being used for road maintenance.

In sum, the system of congestion tolls can play an important role as a source of road financing in the future. Even if all of the above problems are solved, the function of congestion tolls as a source of road financing is still limited to some portions of major intercity highways (those which are near-urban), bridges and major urban roads. It would not play a role in low-density roads including rural roads.

Construction of Toll Roads

In an era of fiscal conservatism, reflected by a shift from public- to private-sector provision of services, increasing constraints on governments' abilities to control resources and a general shift to a greater emphasis on efficiency, construction of toll roads is likely to become an important means of dealing with the increasing need for investment in new road systems. Currently, the toll-road networks are growing rapidly with over 5,400 kilometres in France, in excess of 5,100 kilometres in Italy, 4,700 kilometres in Japan and over 7,100 kilometres in the United States (Heggie, 1991). This can be an increasingly important source of financing for new road projects. Numerous states, including California, Colorado, Virginia and Texas, are making use of or are planning to make use of tolls as an alternative funding source (see Allen and Floyd, 1991). The British Columbia government built the Coquihalla toll highway linking Hope and Kamloops and the Okanagan Valley. Many of these toll roads charge differential tolls between peak and off-peak periods and, therefore, their pricing policy is consistent with the spirit of congestion tolls. An appropriately managed and regulated (including the monopoly nature of toll setting) toll road system can promote efficiency in using a given road capacity in the short run, and the attendant market forces will pressure the road authority into putting optimal capacity in place in the long run. Advances in collection technology are expected to encourage the use of toll facilities for building major intercity highways and urban access roads.

Annual Licence Fees and Taxes Involving the Acquisition of Vehicles

Empirical studies which have investigated the impact of higher licensing fees on peak-hour car driving (raising the fixed cost of car driving relative to common-carrier services) have concluded that such fees have failed to produce any substantial improvement in congestion (Allen and Floyd, 1991). They argue that governments are going to have to substantially raise licensing and registration fees in order to have any impact on reducing car use. A substantial increase in both the one-time registration fee (at the time a new or used vehicle is acquired)³⁷ and annual licence fees from the current level may improve the economic efficiency of using the existing road capacity subject to the condition discussed earlier regarding the elasticity values. If there are substantial fixed road costs which do not vary with traffic volume and the road authority decides to recover these fixed costs from road users (rather than subsidizing them from general tax revenues), it will be more efficient to recover the fixed costs as lump sum annual fees (licence fees) rather than allocating them by marking up the usage-related charges (per-kilometre charge) above the respective marginal costs if the access elasticity is equal to zero. If this is not so, the conditions set out earlier will hold. Specifically, access fees will be reduced and variable charges increased until the net change in welfare with a shift between the two fees is equal to zero; in essence welfare is enhanced by exploiting available gains from trade between the two sources of funds.

Although they are likely to be unpopular, like fuel taxes, these fees are administratively easy to assess and collect. The optimal taxes and charges involving annual ownership and acquisition of vehicles can generate large amounts of revenue for the road sector as these items currently account for 42% of the total charges paid by road users in Japan, 36% in the United Kingdom and 38% in the United States (Heggie, 1991) while accounting for 23% in Canada.

Charging Beneficiaries of the Improved Road Access

For developing local access roads in a suburban area, it is reasonable to require developers to build the roads and to incorporate the costs into the final selling price of the building or service site. This is an economically efficient solution in terms of both the size of the road investment as the developer is expected to make a trade-off between ease of road access and the value of the remaining lands they can sell.

For major improvements in urban road systems, a part of the construction costs may be assessed against the properly identified land-value increments. The concept of special assessment district (under which the geographic areas are identified to levy special assessments), which are being used in several states, is a feasible way to generate funds for large-scale highway reconstruction projects in a reasonably equitable way. This requires close coordination between the government with the taxing authority and the government in charge of road construction. The identification of a special assessment district and actual assessment of special levies can be a controversial process. For example, as soon as the major improvement in the road system is made, the option value of using the road can be realized by the businesses and residents in that area, and, thus, the special levies are economically justified. However, this one-time special levy does not guarantee the efficient use of the road infrastructure, and a combination of one-time special assessment for the construction costs and charging tolls (variable fees) related to road usage would yield a more efficient use of the road and a more efficient level of investment. It may be an interesting option to explore.

5.4 SUMMARY OF ALTERNATIVE ROAD FINANCING METHODS

Several promising sources of road financing are summarized below. Each charging method is evaluated with respect to the promotion of the efficient use of road capacity and durability, the efficiency of (capacity) investment, the potential size of the fund that could be generated, administrative feasibility and collection costs, equity among various groups and the domain of applicability. Several alternative combinations of charging methods are also evaluated and compared.

5.4.1 Summary Evaluation of Financing Methods

Motive Fuel Tax

The evaluation is restricted to its application to car users only. It would be ranked “moderate” in promoting efficient road usage on uncongested roads because the amount of fuel tax one pays is weakly correlated with road damage costs which are the responsibility of the user. It would rank “poor” in promoting the efficient allocation of peak-load capacity since its price is not correlated with road use. It would be regarded as “poor” in promoting investment efficiency on congested roads as the fuel tax does not exert much pressure to build roads with optimal design standards (thickness and

capacity). Administrative ease is a strength as large amounts of funds can be collected easily with relatively small administrative and collection costs. Motive fuel tax can be applied for financing nation-wide and/or state-wide road systems. They are regarded as an equitable method of financing uncongested facilities because every user pays according to the user's consumption of fuel. It would be inequitable on a congested facility since it is not highly correlated with congestion costs.

Graduated Per-Mile Tax Based on Axle Weight

It promotes "reasonably high" efficiency in both pricing road damage and inducing an optimal investment in road durability as this tax can be designed to closely reflect the road damage costs each user imposes, and this will generate enough pressure for the road authority to build roads with optimal thickness. However, if a substantial portion of road damage is due to "environmental factors," this tax is likely to achieve maximum efficiency only if it is combined with a lump-sum tax such as vehicle licence fees. This tax is not intended to be a congestion charge or a charge which would optimally allocate scarce road capacity among competing users. It can generate a large amount of funds if it replaces the current fuel taxes, but it will require some administrative work and collection expense. It is feasible to administer such a tax system. Since users pay costs they impose, it can be considered equitable. To be effective, this system should be applied in all parts of a given jurisdiction such as a province or nationally rather than trying to apply it in only a part of a province, for example. If it is applied consistently on all roads within a jurisdiction, it will induce the efficient use of the road system by trucks as well as the efficient choice of trucking technology given the roadway system. If the tax is applied piecemeal, there will be less incentive for the adoption of trucks which minimize road damage. Higher overall road costs may also result if trucks attempt to avoid the charge on some roads by using non-toll roads and generating even more pavement damage.

Congestion Tolls

Tolls promote high efficiency both in the allocation of scarce road capacity and in inducing the optimal investment in road capacity (number of lanes). Congestion tolls are not intended to be a price for road damage but rather to ensure the roads are used and invested in efficiently. Congestion tolls in major urban or near-urban areas could raise a large amount of money. The Hong Kong experiment has provided convincing evidence that such

a system is feasible but the cost of electronic vehicle identification and collection is expected to be substantial. Some claim that charging congestion tolls is not equitable because "poor working people" pay high congestion tolls and downtown businesses lose out. It is possible to compensate these losers by subsidizing urban transit services from the congestion toll revenues. Once the way in which the toll revenue is spent is taken into account, the actual and perceived inequity of congestion tolls may disappear. The introduction of congestion tolls or peak-load charges should not be undertaken in isolation with the expectation they will provide the solution to the pricing and investment problems. Pricing is one tool of the transportation planner and policy maker. To be effective, it must be used in conjunction with other strategies, such as substitute mode investment. To gain public support for pricing, it is essential that it not be perceived as simply another way of raising revenue for government nor that it is exploiting a situation in which people have no alternatives. For these reasons it is important that alternatives be available for those facing higher fees.

Toll Roads

If they implement peak-period pricing, toll roads have essentially the same characteristics as implementing congestion tolls on the existing roads. Constructing toll roads allows private-sector participation for large project financing. As collection technology improves, toll collection costs become quite moderate. New toll roads benefit not only those who switch to the toll roads but also those who continue to use the old non-toll roads (due to reduced travel time). This is a Pareto improvement. Construction of toll roads can be considered an equitable method since lower-income users gain by travelling on non-toll facilities. In reality, toll roads can solve financing problems for specific projects only.

Annual Licence Fees

These fees can be used to contribute to road costs which do not vary with traffic volume in the short run. If they are not coupled with a variable charge, as with a two-part tariff, which reflects social costs, annual licence fees will have no direct effect on the efficiency of road investment. As shown in many countries, it can generate large sums of money with relatively low administrative and collection costs. Since some people cannot afford to keep a car when annual licence fees are raised substantially, a high licence fee policy may not be regarded as equitable. One method used in some

U.S. states is to base the fee on the purchase price of the car and have the fee decline each successive year to some floor level where the floor would be the same for everyone. A system of licence fees can be implemented by national, provincial or municipal governments.

Charging Beneficiaries

This one-time charge to beneficiaries (developers, property owners and businesses in a special assessment district) is likely to promote an efficient level of road capacity investment, but has no effect on the efficiency of road usage (use of road capacity or durability) after construction. This is a good source of funds for a well-defined road project but it is difficult to determine the boundaries of a special assessment district, as well as the amount of special levies. Charging beneficiaries is by definition equitable. This method of financing works only on a project basis.

5.4.2 Evaluation of Alternative Sets of Financing Instruments

Government must temper principle with practicality and use a mix of two or more charging methods to improve economic efficiency in the road transportation sector while simultaneously attaining a cost-recovery target. Furthermore, perceived equity cannot be ignored in the selection of the menu of instruments. The following alternative sets of financing instruments are evaluated and compared:

- fuel taxes and licence fees (status quo);
- fuel taxes, licence fees and congestion tolls;
- graduated per-kilometre taxes, congestion tolls and licence fees; and
- social marginal cost pricing (graduated per-kilometre taxes based on axle weights, congestion tolls and other externality taxes).

Fuel Taxes and Licence Fees

Canada, as well as most other countries, currently use fuel taxes and licence fees as the major source of funds for financing road infrastructure. These two instruments together can generate total revenue which is sufficient to finance all of the road infrastructure. They are also administratively easy to collect. Licence fees, if appropriately set, can be used to recover some

or all of the short-term fixed cost including the portion of road repair and repayment costs which are unrelated to road usage.

Fuel taxes do not give road users, most importantly heavy trucks and buses, the incentive to choose the optimal axle load which minimizes the combined user and infrastructure costs. This happens because commercial carriers try to reduce fuel consumption which tends to increase with the number of axles, while damage to the pavement increases exponentially with average weight per axle. A second flaw in the use of fuel taxes for pricing urban and near-urban roads is the low correlation between congestion costs and fuel use, hence fuel tax paid. The fuel tax is, therefore, a poor proxy for a congestion tax and would lack any capability for allocating peak-load capacity. Overall, the combination of fuel taxes and licence fees is not an economically efficient method of charging and financing roads.

Fuel Taxes, Licence Fees and Congestion Tolls (Urban and Near-Urban Roads)

The congestion toll component of this package is likely to improve the efficiency of allocating peak-period capacity, and also generate pressure to invest optimally in capacity. However, this package lacks incentives for commercial operators to use optimal axle-load weight.

Graduated Per-Kilometre Taxes, Congestion Tolls and Licence Fees

This combination of fees may be regarded as a transition package en route to a more socially efficient pricing scheme. This study has already addressed the value of fuel taxes and licence fees as providing a practical and effective means of pricing and financing capacity investment for intercity roads on which there is little or no congestion. The weak link is the failure to adequately price truck damage and provide incentives for operators to employ economically efficient axle loads. The fuel tax should therefore be restricted to cars and the scheme supplemented by a charge for trucks.

A graduated per-kilometre axle fee should supplement or replace current truck licence fees and fuel taxes levied on truck use. This will not only efficiently and adequately allocate cost responsibility between cars, trucks and buses but also among the various types of trucks.

Licence fees which are levied will reflect, not necessarily on a one-to-one basis, the magnitude of fixed costs. Licence fees must be allocated both between and within user groups. This means that for fairness and efficiency, trucks, cars and buses should bear a portion of the fixed costs of the roadway system — how much will depend on cost responsibility, the elasticity of access with respect to licence fees and the elasticity of usage with respect to licence fee. User-group allocation would be based on some Ramsey-type approach; for example, car users may have a licence fee based on the purchase price of the vehicle — an approach used in some U.S. states. This study does not advocate the position that the contribution by trucks and buses must be contained in a licence fee. Indeed, it may well be that the graduated per-kilometre usage fee will be the only fee levied and contain both usage and fixed contributions.

For urban and near-urban roads this pricing approach will prove to be neither equitable nor efficient. Some method of congestion pricing must be introduced to efficiently allocate existing capacity and provide for efficient investment in future capacity. Fairness, however, will be determined by the expenditure of the congestion tax revenue and taking a comprehensive approach. That is, it serves no one to focus entirely on one instrument and rely on it to provide a solution. A comprehensive approach is necessary, in which substitutes are provided for those faced with a congestion or peak-load fee. These substitutes can range from flexibility of work hours to providing public transit (of some form).

Social Marginal Cost Pricing

The graduated per-kilometre taxes based on axle weights reflect the cost of road damage users inflict, while congestion tolls and other externality (noise and air pollution) taxes reflect the costs users impose on the community. Together they constitute the social marginal cost of using the road system, and charging the social marginal costs maximizes social welfare. Although the results of Small, Winston and Evans (1989) suggest that this charging scheme would result in an approximate financial break even, based on the assumption of constant returns to scale in joint production of road capacity and durability, other evidence such as Nix (1989), Nix, Boucher and Hutchinson (1992), Kraus (1982) and Oum and Zhang (1990) suggests that break-even performance is not guaranteed. In Canada where the majority of

the non-urban roads are of low density (and capacity) and the indivisibility of capacity construction plays some role, the application of social marginal cost pricing is likely to produce a financial deficit in many instances.

As we have also discussed, there are also some practical and administrative difficulties in implementing the graduated per-kilometre charges based on axle weight, and the externality taxes (congestion tolls, noise and air pollution tax). Certainly, it will result in higher costs for administration and collection than the current fuel tax and licence fee combination. However, this is the correct direction in which to move for maximizing economic welfare from the road sector. As the technology of automatic vehicle identification develops and the cost of information processing decreases, the administration and collection is likely to become less burdensome than now.

6. SUMMARY AND CONCLUSIONS

6.1 PURPOSE AND ORGANIZATION OF THE STUDY

Transportation policy has undergone significant changes over the last 50 years. Canada has moved from a situation of government ownership, management and regulation in all or parts of every mode to a position of deregulation, privatization and a generally greater reliance on market forces. The major transformation has taken place among the carriers while more recent initiatives have been focussing on infrastructure — roads, airports, rail track and terminals.

Much of the stimulus for change has been the result of Canada's need to rethink traditional policies and reposition itself to remain competitive in the North American and world marketplaces. Pricing and investment planning of Canada's transportation infrastructure cannot ignore the forces and pressures developing in the United States and other international markets. Links must be established between investment in infrastructure and the pricing of services delivered by that infrastructure. Indeed, socially optimal modal pricing requires the inclusion of modal air and noise pollution and congestion externalities. Economic welfare will be lower if these factors are not considered in modal prices because demands for massive public investments in infrastructure will continue unabated, infrastructure will deteriorate prematurely, and the distribution of traffic across modes will not reflect the

real costs of these modes. Solutions do not necessarily lie with more investment but rather with smarter investment. Smarter investment must start with efficient pricing.

In May 1992, the Congressional Budget Office of the United States released a study, *Paying for Highways, Airways and Waterways: How Can Users be Charged?* The study is in complete congruity with the conclusions stated here. They state in their summary:

The methods of financing highways, airways, and waterways influence both the amount of revenue that can be raised and the efficient allocation of resources. The concept of revenue adequacy — whether revenues cover costs — is important to the cash-strapped federal government, but it also has implications for efficient allocation of resources in the long run. If the costs of an investment project cannot be recovered from those who use it, the project's feasibility comes into question. But an investment that benefits society is worth making, even though it may not be possible to charge users for it. This often characterizes goods and services provided by the federal government, and it underlies the rationale for government rather than private activity in certain sectors. Revenue adequacy can provide information about the demand by users for public investments, but it alone cannot be the criterion upon which investment decisions are made.

Economic efficiency is the second criterion by which financing mechanisms are evaluated. The standard definition of allocative efficiency is used here: does the price — the value consumers place on the product or service at the margin — equal the marginal cost — that is, the value of resources used in producing the last unit? If the price is less than the marginal cost, consumers tend to overuse the resource; if the price exceeds the marginal cost, they use it too little.

The objectives of revenue adequacy and economic efficiency sometimes conflict. Economic theory offers some ways of minimizing the trade-offs, and these are included in the discussions of alternative pricing mechanisms. (p. xi)

The main purpose of this study is to discuss the principles and methods by which Canada can ensure the optimal use and efficient provision of transportation infrastructure services. Also, to reflect the move to greater fiscal

responsibility reflected in a shift from public- to private-sector provision of services, increasing constraints on government's ability to control resources and a general shift to a greater emphasis on efficiency, an emphasis is placed on the cost-recovery issue. The issue is treated in two alternative ways. First, the cost-recovery conditions associated with optimal pricing and investment in infrastructure are studied and compared with the actual cost-recovery situation. Second, the methods of achieving an exogenously given cost-recovery target in such a way as to minimize the efficiency loss are discussed. Although the principles and methods discussed in the study can be applied to all modes of transportation, the emphasis is on roads and airport infrastructure.

6.2 INFRASTRUCTURE PRICING PRINCIPLES

Pricing is a method of allocating resources. There is no such thing as the right price irrespective of issues and objectives. Rather there are optimal prices or pricing strategies given particular objectives to be achieved. Prices can be established to maximize profit, welfare or revenues. They can be used to achieve a particular market share or a desired distribution of demand across products (for example, mode-split in transportation). However, one of the most important goals for pricing goods and services from society's viewpoint is in maximizing economic welfare by optimally allocating scarce resources and goods and services across competing needs in the short run and ensuring optimal investment in capacity in the long run.

It is well known that economic welfare is maximized by pricing infrastructure at marginal social cost which means the externality costs users impose on the system are included. Critics of the current pricing approach agree that governments have been misdirected in their infrastructure policy chiefly in trying to respond to traffic growth by expanding capacity rather than also using demand management through an efficient pricing system. Empirical studies indicate that the net social benefit from effective demand management policy is quite high. The major source of the available efficiency gains for both near-urban roads and airports is the reduction in congestion delays with a consequent savings in operating and time costs. For intercity facilities, efficient pricing leads to a more efficient use of facilities within and across modes.

Efficient user-charges which reflect social costs have three significant impacts. First, economic welfare is increased as the demand for infrastructure is rationed more efficiently in a way which reflects the costs of using the

infrastructure. Second, the distortion in traffic allocation across modes is reduced. Third, it contributes to the financing of infrastructure. Furthermore, marginal social cost pricing is not incompatible with fairness or equity. The surplus from congestion tolls may be used to offset losses by some groups who may be adversely affected by increased user-fees. The British Airports Authority (BAA Plc.), for example, has used marginal cost pricing principles to establish runway and terminal charges for the five airports under its control but has redistributed revenues from one airport to another particularly for investment purposes.

Peak-load pricing is an important variant of marginal social cost pricing which formally considers the fact that some users demand and require more capacity than others. This is seen everyday on roadways, runways and in terminals. There is a clear and sound economic efficiency basis for peak-load pricing. Charging higher prices to peak users enhances economic efficiency by inducing them to make rational choice decisions as well as helping to solve financing problems for capacity expansion. Finally, peak pricing is not, in principle, unfair or inequitable. It assigns costs to those who are responsible for them. It makes no economic sense to restrict the use of a facility in off-peak hours because all that results in is the underutilization of an existing facility. In competitive circumstances or under regulation, peak users are not necessarily worse off in terms of what they pay, if there is a lower off-peak price, provided the off-peak users pay at least their variable costs. If off-peak users were eliminated from the market, peak users would pay precisely the same amount as when off-peak users pay their variable costs. The only circumstance when peak users would be better off is if they are cross subsidized from off-peak users. This creates a social loss in terms of both inefficient use of facilities and the loss of consumer surplus.

When the total revenue from the social marginal cost pricing is not sufficient to cover total cost, there are three options open to the infrastructure authority:

- continue to use social marginal cost pricing with subsidization from the general tax revenue;
- use Ramsey pricing which provides a second-best solution with a break-even financial performance; or
- adopt a two-part pricing (access/usage tariffs) scheme.

The question of financing capacity from users versus the general revenue fund (all taxpayers) is an issue of both efficiency and equity. If marginal cost is less than average cost, efficient pricing will result in a deficit. This deficit must be covered. A partial equilibrium approach to this problem would support a policy of funding the deficit out of general revenue while a general equilibrium approach would argue that the relative welfare losses of funding the deficit by users versus the general taxpayer should be considered. Recent evidence, Jorgenson (1992), Jorgenson and Yun (1990) and Ballard, Shoven and Whalley (1985), has shown that the marginal cost of public funds is between \$1.33 and \$1.45, meaning that between 33¢ and 45¢ per dollar are lost through a private-sector efficiency loss. It is, therefore, not obvious that proposing a policy of marginal social cost pricing and ignoring the costs of public funds to finance the deficit would necessarily improve economic welfare. A second-best pricing scheme in which there is a cost recovery constraint may lead to a higher level of economic efficiency.

Ramsey pricing minimizes the loss of economic efficiency caused by deviating prices from the respective marginal costs in order to achieve financial break even. In effect, it charges higher markup to less price-elastic products or market segments by making the markup inversely proportional to the price elasticity of the demand.

Two-part pricing consists of a flat fee for the right to access a facility (for example, vehicle licence fee to access the road system), and a usage fee (for example, charge per kilometre of road usage). Two-part pricing can lead to a first-best solution if the demand for access is not price sensitive, and if usage charges are set at the marginal costs of usage and the access fee is set at a sufficiently high level to allow the firm to break even. Under these conditions, the regulation of the access fee alone can induce a monopoly firm to charge marginal cost as the usage price. When access demand is sensitive to price, the optimal two-part tariff can be computed by applying the Ramsey pricing rule to the access and the usage demands as if they are two separate products with interrelated demands. This then becomes a second-best pricing approach.

The economics literature is paying increasing attention to the consequences of selective pricing methods for various groups of users and non-users; that is, the income distributional consequences of a pricing method. Problems arise when a pricing method which maximizes economic efficiency does not necessarily yield what is considered a fair or equitable outcome. The

most obvious case where concern over income distribution appears critical is the implementation of congestion (or peak-load) pricing for (urban) transportation. The results of a number of empirical and simulation studies vary somewhat, but the basic conclusion is that it is misleading to characterize a congestion or peak-load pricing policy as being regressive.

Marginal social cost pricing is also not incompatible with fairness or equity. Congestion or peak-load pricing in conjunction with a strategy to use the revenues can generate net positive benefits for society. Marginal social cost pricing corrects distortions rather than introduces them. Some groups, however, are made better off and others worse off but this should not justify rejection of this pricing policy. As Hau (1991) notes, it is perhaps asking too much of a pricing mechanism to solve the pricing, investment and income distribution problem. One perspective is to use road and airport funds from efficient pricing to both invest efficiently in capacity (new roads or runways) and to satisfy concerns regarding equity. This study recognizes that while income redistribution issues are better placed in general tax policy, some of the concerns with the perceived inequity of efficient pricing may be allayed if some of the funds from efficient pricing are used to provide substitutes for those most affected by the use of socially efficient pricing. Thus, a transportation fund could potentially invest in public transportation, rural roads or small airports. The magnitude of the fund would depend on scale economies and indivisibilities.

6.3 CARRIER AND INFRASTRUCTURE COSTS

The infrastructure planner must establish user-charges and make capacity investment decisions to maximize the economic welfare of society. Understanding the behaviour of the combined cost of a carrier's service provision and infrastructure provision is essential for the development of a set of socially optimal prices for infrastructure. If short-run costs fall because of increased capacity utilization but long-run costs exhibit constant returns to scale, it is still possible to have marginal social cost pricing and fully cover costs. If, however, long-run costs are characterized by some economies, a second-best pricing approach will be required to have total revenues cover costs and minimize the efficiency loss of deviating from first-best pricing. A prerequisite to understanding the structure of the combined cost is to understand each component, that is, carrier cost structure and infrastructure cost structure.

A number of studies have been directed at determining the behaviour of an airline's cost function with respect to changes in the level and composition of output. The studies have shown that the long-run average cost curve is relatively constant over a wide range of output, that is, there are no economies of scale in the airline industry. This means that the size of a carrier does not generate lower per-unit costs. Studies also concluded that airport capacity construction is as well characterized by constant returns to scale. This implies that the combined cost of carriers and infrastructure is also characterized by constant returns to scale.

Empirical investigations of the truck and intercity bus industries have provided mixed results. The overwhelming evidence is that there are constant returns to scale in trucking, some scope economies and certainly density economies. The limited investigations of the bus industry found weak evidence of non-constant returns. There is no direct evidence of density economies but observing the restructuring of the bus industry in those countries in which there has been deregulation certainly suggests some density economies. Evidence from Small, Winston and Evans (1989) showed constant returns to scale in highway capacity construction. They reported the existence of significant economies of scale with respect to the durability of road and mild returns to scale with respect to traffic volume. They also reported diseconomies of scope from the production of both durability and traffic volume because, as the road is made wider to accommodate more traffic, the cost of any additional thickness rises, since all the lanes must be built to the same standard of thickness.

The final outcome of these three factors at work is that highway capacity construction is characterized by approximately constant returns to scale. In other words, the output-specific economies of scale are offset by the diseconomies of scope for having to produce them jointly. Since they included both infrastructure costs and the costs incurred by road users (individual drivers and transportation carriers) in the total cost of highway modes, their result on the overall constant returns to scale is for the combined cost of highways and users.

6.4 IMPLEMENTATION OF USER-CHARGES

This review of the current charging schemes for roadways and airports and of efficient pricing has led to the following conclusions. In the air sector the uniform user-charge system should be repealed in favour of site-specific

user-charges. Furthermore, the current weight-based landing fees are too low for small aircraft (mainly general aviation and corporate aircraft) and too high for large aircraft. This had led to a situation of undercharging the former relative to commercial aircraft (mainly large jet).

Peak-period user-charges should be implemented at airports with congestion problems, and the differential user-charges between peak, low-peak, and off-peak periods can be set to reflect their respective social marginal costs which include congestion externality costs. This leads to a large differential between peak and off-peak fees. Under this system, landing fees for small aircraft become similar to those for large aircraft in the peak periods. It is clear that landing fees for small aircraft at congested airports are even more underpriced. Moving in this direction is highly consistent with the pricing provisions contained in the Transport Canada cost-recovery proposals published in 1990.

A weight-based landing fee, if appropriately set, may be consistent with the Ramsey pricing principle and, thus, have some economic justification for use at small uncongested airports. This is because the demand for airport services by larger (heavier) aircraft is more price-inelastic than smaller (lighter) aircraft. However, the rationale for weight-based pricing breaks down when the airport becomes congested and congestion pricing is applied. Similarly, charging higher landing fees for long-haul international flights at uncongested airports can be justified in an efficient pricing regime since it is consistent with the spirit of Ramsey pricing as the demand for airport services by long-haul flights is less price-elastic than short-haul flights. However, charging higher landing fees for long-haul international flights loses its economic rationale when congestion pricing is applied to congested airports.

In establishing prices for roadways, the empirical studies conducted in Canada and the United States indicate passenger cars and other light vehicles pay a disproportionately higher share of the total road maintenance costs as compared to trucks and other heavy vehicles. The Canadian studies show that the short-run fixed road cost far exceeds the user-charge revenues not related to road usage. This implies that vehicle licence fees may be too low to be optimal, and those who underutilize their vehicles pay less for their option to use the road system than the optimal price for the option to use their vehicle.

Road costs are sensitive to the investment level in capacity and durability. It is therefore important that the road authority make an optimal investment particularly in durability (pavement thickness). In the United States, the underinvestment in durability (that is, building sub-standard roads) has not only substantially increased road maintenance costs more than necessary, but also the costs to the road users.

Currently, the majority of the road user-charges are collected in the form of fuel taxes. Since fuel taxes are not directly related to either congestion or road damage, they are not an efficient means of charging for road usage. Road charges must be related to both road damage and the amount of congestion. Technology now exists for effectively administering congestion tolls either by time of day or by congested road segment. Therefore, the users of urban roads and near-urban highways must be charged both congestion tolls and road damage while users of rural and uncongested intercity roads are charged just for the road damage.

6.5 ALTERNATIVE SOURCES OF FINANCING ROAD INFRASTRUCTURE

In practice, a government must use a mix of two or more charging methods to achieve high efficiency in the road transportation sector while attaining a cost recovery target. The road pricing scheme recommended by this study consists of three parts:

- a system of graduated per-kilometre fees for trucks based on axle weight which will improve efficiency in the use of roads and promote an optimal investment in the durability (thickness) of the road system;
- congestion tolls and environmental externality taxes which will help improve the efficiency of the usage of road capacity for urban and near-urban roads as well as achieve an optimal investment in road capacity (lanes); and
- a fuel tax–licence fee combination for cars on uncongested roads.

The relative mix of fixed and variable fees will depend on the relative values of access and usage price elasticities. It will generally be true that a greater proportion of revenues will be generated from the variable charge.

The first two are essentially the social marginal costs of road usage (that is, charging for road damage and externality costs). Construction of toll roads or charging beneficiaries are essentially local financing solutions for specific road construction or improvement projects. Since empirical evidence suggests that a substantial portion of road costs do not vary with road usage (traffic volume or axle loads), use of substantial licence/registration fees is likely to be an efficient means of recovering the fixed road costs if the number of users is not affected by the magnitude of the licence fee. Methodologically, a two-part pricing system can be applied on a national or provincial scale to charge vehicle licence fees for the right to access the road system and social marginal costs (road damage and externality costs) for the road usage. This two-part pricing system is likely to self-finance the Canadian road system.

This study recognizes that motive fuel taxes have been deeply entrenched in society because it is administratively easy to collect large sums of money from road users but also because it helps achieve other important objectives such as environmental and energy conservation goals. However, when planning for 20 or 30 years into the future, the user-charge plans must include economically rational charging methods such as peak-period and peak-district pricing and graduated per-kilometre charges based on axle loadings if there is to be an efficient system.

ENDNOTES

The authors acknowledge the intellectual contribution of Yimin Zhang, and the research assistance of Eva Lau and Nagy Eltony. We are indebted to Richard Arnott, Sandy Borins and particularly John Sargent for helpful comments and suggestions on substance and exposition. We, however, are responsible for the contents of the document.

1. For the case of a single output with no possibility of effective market segmentation, Ramsey pricing reduces to average cost pricing.
2. Oum and Tretheway (1988) have extended the Ramsey pricing rule to the case where externality costs are present. Their results show that the markup is a weighted average of the inverse elasticity and the ratio of marginal external cost to price.
3. Essentially the same method as Ramsey pricing can be applied to find quasi-optimal prices to achieve a given level of cost recovery (including break even). See Gillen, Oum and Tretheway (1986) for use of the Ramsey framework to compute optimal landing fees subject to various levels of cost recovery for runway services. In the past, some transport economists named it "value-of-service" pricing when firms (airlines and railroads) charged differential (often profit-maximizing) prices for essentially the same services by using differential price elasticities by market segment.

4. The following discussion is in the context of two-part prices but the concepts are not limited to two parts and can apply equally well to multi-part prices.
5. There are a number of examples of two-part prices in the economy: pricing telephone services, golf and other club memberships, amusement parks, roadway access and usage.
6. Train (1991) notes that too high an access fee may result in hardship for lower income groups. If this is perceived as being too inequitable, the efficiency gains may be foregone.
7. See *Aviation Week and Space Technology*, April 27, 1992.
8. Ng and Weisser (1974) have shown that the proportion of the "residual revenue" (the difference between revenue required to break even and the revenue obtained from pure marginal cost pricing) financed by increasing the variable fee above marginal cost, increases with the absolute price elasticity of the number of users and decreases with the absolute price elasticity of the amount of usage.
9. See Ng and Weisser (1974), Spulber (1989) and Train (1991) for further discussions on theoretical and technical issues on multi-part pricing, and its relationship with "block tariffs."
10. Investment to increase slots does not always require a new runway. Construction of a high-speed turnoff, for example, can decrease the amount of time a particular plane occupies a runway.
11. In fact, the landing fee alone is identical in the peak periods for all aircraft types (see Gillen, Oum and Tretheway, 1988).
12. There are numerous studies of urban road pricing and its consequences but they are not reviewed here since the interest is in interurban infrastructure.
13. In all cases the evaluations or comparisons are made against a status quo which is essentially average cost pricing. Thus, the comparison is between average costs and social marginal cost pricing for congested facilities and between average cost and Ramsey pricing for uncongested facilities.
14. Small, Winston and Evans (1989) also developed a set of efficient road prices. These are discussed at greater length later in the section dealing with the question of cost recovery.
15. This was changed in 1991.
16. Since the total disbursements on roads in the United States was \$61 billion in 1985 (Small, Winston and Evans, 1989), this means a savings of over 13 percent of the total cost.
17. A recent study in Minneapolis-St. Paul found that the addition of one runway to the international airport would generate such large savings in direct cost to carriers that it (the runway) would have a payback of only three years.
18. Underpricing infrastructure not only leads to excessive use but oversupplying infrastructure leads to long-term structural shifts which almost guarantees today's problems will recur in the future.
19. See Keeler and Small (1977) and Winston (1985) for examples of modelling optimal user charges and investment decisions for infrastructure in such a way as to maximize total benefits for users (social welfare).

20. In the case of a multi-product firm, outputs would be increased in the same proportion or along an output ray.
21. As a corollary, if there are no scale economies, it implies that first-best pricing with no deficit is achievable over the long run.
22. The exception is Air Canada which Gillen et al. (1985) found to have realized most of the available density economies in its network. This may have changed since the empirical evidence is based on data up to 1981 only.
23. Since Canada has a number of small airports in remote communities, indivisibility of capacity at small airports may translate into scale economies.
24. This is one of the significant problems associated with system average costing. It assumes there are not significant differences in costs and demand between facilities.
25. The numbers are also based on costs at the time the study was conducted. The numbers calculated would be as high or lower since lower costs are expected with the move to defederalization (see Hamilton, 1991).
26. Airport planning and construction can in many cases take more than a decade so the characteristics of commercial aircraft soon to be produced are considered in the design.
27. Table 4.1 also shows differential prices by market segment: domestic and international flights. This may be consistent with the spirit of the Ramsey pricing principle considering that most international flights have longer stage lengths and, thus, lower price elasticities of demand for airport services.
28. GA aircraft did pay an airport-specific tax to cover the costs of airport operations but this was not an explicit fee for runway use.
29. As an example of this type of charging structure, New Zealand which has privatized its air traffic control system, charges general aviation aircraft \$57 per year (plus sales tax) for the first 50 landings and \$3.67 or \$4.60 for each subsequent landing depending on airport location (see Paul Proctor, "For Profit New Zealand ATC System Cuts Costs and Increases Efficiency," *Aviation Week and Space Technology*, April 27, 1992).
30. Many of these changes are consistent with a recent Transport Canada pricing policy proposal (Transport Canada, 1990).
31. For example, at Heathrow airport all aircraft, small or large, pay identical landing fees during peak periods. In addition, the combination of landing fees and passenger terminal fees for a large aircraft (for example, Boeing 747) in the peak periods exceeds, by a factor of five, the amount for the same aircraft at off-peak periods. See Gillen, Oum and Tretheway (1988) for the exact fee differentials.
32. Nix allocated costs under two scenarios: with scenario A all capital costs are treated as inescapable (assumed not to vary with vehicle usage). With scenario B one third of the capital costs are assumed to be escapable. Among the capital costs, the pavement costs are allocated to various axle-weight groups, while the road maintenance costs (and one third of the pavement costs in scenario B) were allocated on the basis of vehicle usage. This cost allocation exercise was performed on the following vehicle categories: standard car, three-axle truck, five-axle tractor-semitrailer, six-axle tractor-semitrailer, eight-axle B-train loaded and eight-axle B-train empty.

33. Heggie (1991) categorized the charging instruments into four categories: vehicle usage, vehicle ownership, vehicle acquisition and charging beneficiaries of road system.
34. One exception to this is the case of Japan where substantial revenue from road tolls is generated. Toll revenues accounted for nearly 20 percent of the total revenue collected from road users in 1985. In fact, all three publicly owned road authorities generate financial surpluses after covering capital expansion costs.
35. Small, Winston and Evans (1989) proposed the use of both a steeply graduated per-mile tax based on axle weights and congestion tolls as the main feature of road pricing (pp. 114-19).
36. The Hong Kong experiment is interesting, in part because it is sometimes cited as an example of the failure of congestion charges. Hong Kong did not adopt the congestion toll system after the experiment carried out from 1983 to 1985 primarily because of the complex political factors including people's fears of government intrusion and the desire to exercise the newly won local autonomy from the British governor. However, the test of collection technology in Hong Kong was a resounding success, exceeding by a wide margin the very stringent goals for reliability and ease of use that were established. More than 99.7 percent of vehicles crossing toll sites were correctly identified, and the wrong vehicle was charged less than one time in 10 million.
37. This one-time registration fee may have a small short-run negative effect on the environment since it would discourage the purchase of new vehicles and encourage continued use of older ones. Therefore, it may be desirable to raise annual licence fees substantially without changing the first-time vehicle registration fees.

REFERENCES

- Abouchar, A. 1970. "Air Transport Demand Congestion Costs and the Theory of Optimal Airport Use." *Canadian Journal of Economics* 3, pp. 463-75.
- Action, J. P. and I. Vogelsang. 1989. "Symposium on Price-Cap Regulation." *Rand Journal of Economics* 20, no. 3, Autumn, pp. 368-472.
- Aigner, D. and L. Lillard. 1982. "Measuring Peak-Load Pricing Response for Experimental Data." Working Paper 8224, Social System Research Institute, University of Wisconsin, Madison.
- Allen, T. M. and Charles F. Floyd. 1991. "Alternative Financing Techniques in Funding Major Highway Reconstruction Projects." *Transportation Quarterly* 45 (3), July, pp. 357-68.

- Anderson, F. J. and N. C. Bonsor. 1974. "Pricing and Valuation of Transportation Facilities in the Presence of Congestion." *Economica*, November, pp. 424-31.
- Armstrong-Wright, A. T. 1986. "Road Pricing and User Restraint: Opportunities in Developing Countries." *Transportation Research* 20A, pp. 123-27.
- Arnott, R. and J. Stiglitz. 1989. "Congestion Pricing to Improve Air Travel Safety." In *Transportation Safety in an Age of Deregulation*. Edited by Leon Moses and Ian Savage. Oxford University Press.
- Arnott, R., Andre de Palma and R. Lindsey. 1990. "Economics of a Bottleneck." *Journal of Urban Economics* 27, pp. 111-30.
- Ballard, C., J. Shoven and J. Whalley. 1985. "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States." *American Economic Review* 75, pp. 128-39.
- Baumol, A. W. and D. Bradford. 1970. "Optimal Departures from Marginal Cost Pricing." *American Economic Review* 60, pp. 265-83.
- Bisson, B. G. and E. Hildebrand. 1989. "Highway Finance — A Literature Review." Mimeo. Fredericton: The University of New Brunswick Transportation Group.
- Boardman, A. E. and Lester Lave. 1977. "Highway Congestion and Congestion Load Pricing." In *Marginal Cost Pricing in Practice*. Edited by James R. Nelson. Englewood Cliffs, N.J.: Prentice Hall.
- Boiteux, M. 1960. "Peak Load Pricing," *Journal of Business* 33, April, pp. 157-59.
- Bonsor, N. 1984. *Transportation Economics: Theory and Canadian Policy*. Toronto: Butterworths Publishers.
- Borins, S. 1978. "Pricing and Investment in a Transportation Network: The Case of Toronto Airport." *Canadian Journal of Economics* 11, no. 4.

- 1982. "The Economic Effects of Non-Optimal Pricing and Investment Policies for Transportation Facilities." *Transportation Research* 16B, no. 1, pp. 17–29.
- 1984. The Economics of Non-Optimal Pricing and Investment Policies for Substitutable Transportation Facilities." *Canadian Journal of Economics* 17.
- Braeutigam, R. B. 1979. "Optimal Pricing with Intermodal Competition." *American Economic Review* 69, no. 1, March, pp. 38–49.
- 1985. "Efficient Pricing with Rivalry between a Railway and a Pipeline." In *Analytical Studies in Transport Economics*. Edited by Andrew F. Daughety. New York: Cambridge University Press.
- Braid, R. M. 1989. "Uniform versus Peak-Load Pricing of a Bottleneck with Elastic Demand." *Journal of Urban Economics* 26, pp. 320–27.
- Bryan, I. A. and Y. Kotowitz. 1978. *Shipping Conferences in Canada*. Consumer and Corporate Affairs Canada, Research Monograph number 2, Research Branch, Bureau of Competition Policy. Ottawa: Supply and Services Canada.
- Butcher, E. W. A. 1990. "Inter-State Commission Proposed National Road Pricing and Funding Scheme: A Review." *Australia General Public Service*. July, Canberra.
- Button, K. J. 1982. *Transport Economics*. London: Gower Publishing Co.
- Canadian Tax Foundation. *Provincial and Municipal Finances*. Various issues, Toronto.
- *The National Finances 1990*. Toronto.
- Carlin, A. and R. Park. 1970. "Marginal Cost Pricing of Airport Runway Capacity." *American Economic Review* 60, pp. 310–19.
- Carll, R. 1975. "Some Observations on Urban Transport Pricing and its Relationship to Comprehensive Transportation Planning." *Transportation Journal* 14, no. 3 (Spring), pp. 18–30.

- Catling, Ian and B. Harbord. 1985. "Electronic Road Pricing in Hong Kong: The Technology." *Traffic Engineering and Control* 26, pp. 608–15.
- Caves, D., L. R. Christensen and J. Swanson. 1981. "Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955–1974." *American Economic Review* 71 (December), pp. 994–1002.
- Caves, D., L. R. Christensen and M. W. Tretheway. 1980. "Flexible Cost Functions for Multi-product Firm." *Review of Economics and Statistics*, August, pp. 477–81.
- 1984. "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ." *Rand Journal of Economics* 15, no. 4, Winter, pp. 471–89.
- Caves, D., L. R. Christensen, M. W. Tretheway and R. Windle. 1985. "Network Effects and the Measurement of Returns to Scale and Density for U.S. Railroads." In *Analytical Studies in Transport Economics*. Edited by Andrew F. Daughety. New York: Cambridge University Press.
- Charney, A., N. Sidhu and J. Due. 1977. "Short Run Cost Functions for Class II Railroads." *Logistics and Transportation Review* 17, pp. 345–59.
- Clark, J. M. 1923. *Studies in the Economics of Overhead Costs*. Chicago: University of Chicago Press.
- Cohen, Y. 1987. "Commuter Welfare Under Peak-Period Congestion Tolls: Who Gains and Who Loses." *International Journal of Transport Economics* 14, pp. 239–66.
- Cox, W. and J. Love. 1991. "International Experience in Competitive Tendering." A Paper Presented at the 3rd International Conference on Privatization and Deregulation in Passenger Transportation. Tampere, Finland, June 16–20, 1991.
- Damus, S. 1981. "Two-part Tariffs and Optimal Taxation: the Case of Railroad Rates." *American Economic Review* 71, no. 1 (March), pp. 65–79.
- 1984. "Ramsey Pricing by U.S. Railroads." *Journal of Transport Economics and Policy*, January, pp. 51–61.

- Daughety, A. F. 1985. "Transportation Research on Pricing and Regulation: Overview and Suggestions for Future Research." *Transportation Research* 19A, no. 5, pp. 471–87.
- Davies, J. E. 1989. "Cost Recovery Arrangements for Interstate Land Transport in Australia: A Review of Recent Policy Initiatives." *Economic Analysis and Policy*, March 19, no. 1, pp. 91–102.
- Dawson J. A. L. and F. N. Brown. 1985. "Electronic Road Pricing in Hong Kong." *Traffic Engineering and Control* 26, no. 11, pp. 522–29.
- de Borger, B. and W. Nonneman. 1981. "Statistical Cost Functions for Dry Bulk Carriers." *Journal of Transport Economics and Policy* 15, pp. 155–65.
- De Palma, A. and Richard Arnott. 1986. "Usage-Dependent Peak-Load Pricing." *Economic Letters* 20, pp. 101–5.
- DeVany, A. and T. R. Saving. 1980. "Competition and Highway Pricing for Stochastic Traffic." *Journal of Business* 53, pp. 45–60.
- Deweese, D. 1979. "Estimating the Time Costs of Highway Congestion." *Econometrica* 47, pp. 1499–1512.
- Dixit, A., J. Mirrlees and N. Stern. 1975. "Optimal Savings with Economies of Scale." *Review of Economic Studies* 42, pp. 131–46.
- Douglas, G. and J. Miller. 1974. *Economic Regulation of Domestic Air Transport: Theory and Policy*. Washington, D.C.: Brookings Institution.
- D'Ouille, E. L. and John MacDonald. 1990. "Optimal Road Capacity with a Suboptimal Congestion Toll." *Journal of Urban Economics* 28, pp. 34–49.
- Dreze, Jacques. 1964. "Some Postwar Contributions of French Economists to Theory and Public Policy, with a Special Emphasis on Resource Allocation." *American Economic Review* 54, (Supplement).
- Dupuit, J. 1849. "On Tolls and Transport Charges." Reprinted in *International Economic Papers*, 1962, pp. 7–31.

- Foster, C. D. 1974. "The Regressiveness of Road Pricing." *International Journal of Transport Economics* 1, pp. 133–41.
- 1975. "A Note on the Distributional Effects of Road Pricing: A Comment." *Journal of Transport Economics and Policy* 9, pp. 186–88.
- Friedlaender, A. F. 1969. *The Dilemma of Freight Transport Regulation*. Washington, D.C.: Brookings Institution.
- Friedlaender, A. F. and R. H. Spady. 1981. *Freight Transport Regulation: Equity, Efficiency and Competition in the Rail and Trucking Industries*. Cambridge, Mass.: MIT Press.
- Friedlaender, A. F. and S. Mathur. 1982. "Price Distortions, Financing Constraints and Second-Best Investment Rules in the Transportation Industries." *Journal of Public Economics* 18, pp. 195–215.
- Friedlaender, A. F. and S. J. W. Chiang. 1984. "Output Aggregation, Network Effects and the Measurement of Trucking Technology." *The Review of Economics and Statistics* 66, no. 2, (May), pp. 267–76.
- Friedlaender, A. F. and S. S. Bruce. 1985. "Augmentation Effects and Technical Change in the Regulated Trucking Industry 1974–79." In *Analytical Studies in Transport Economics*. Edited by A. F. Daughety, New York: Cambridge University Press.
- Gillen, D. W. and T. H. Oum. 1984. "A Study of Cost Structures of The Canadian Intercity Motor Coach Industry." *Canadian Journal of Economics* 17, no. 2, (May), pp. 369–85.
- Gillen, D. W., T. H. Oum, and M. W. Tretheway. 1985. *Airline Cost and Performance: Implications for Public and Industry Policies*. Vancouver, B.C.: University of British Columbia, Centre for Transportation Studies.
- 1986. *Pricing Policies for Canadian Airports with an Emphasis on Airfield Operations*, a report prepared for the Airports Task Force, Transport Canada, Ottawa.

- 1987. *Measurement of Social Costs: An Application to Toronto International Airport*. Vancouver, B.C.: University of British Columbia, Centre for Transportation Studies.
 - 1988. *A Study of Peak Period Pricing with an Application to Toronto International Airport*. Report written for Airport Authority Group, Transport Canada, Ottawa.
 - 1989. "Airport Pricing Principles: An Application to Canadian Airports." *Journal of the Transportation Research Forum* 29, no. 1, pp. 28–34.
 - 1990a. "Airline Cost Structure and Policy Implications: A Multi-Product Approach for Canadian Airlines." *Journal of Transport Economics and Policy*, January, pp. 9–33.
 - 1990b. "Airport Pricing and Capacity Expansion: Economic Evaluation of the Alternatives." In *Transportation in Canada*. Transport Canada TP10451-E. Ottawa: Transport Canada, Policy and Coordination, Economic Analysis, Chapter 12, pp. 83–93.
- Glaister, Stephen. 1974. "Generalized Consumer Surplus and Public Transport Pricing." *Economic Journal* 84, pp. 849–67.
- 1981. *Fundamentals of Transport Economics*. New York: St. Martins Press.
- Glaister, S. and D. Lewis. 1978. "An Integrated Fares Policy for Transport in London." *Journal of Public Economics* 9, pp. 341–55.
- Gomez-Ibanez, J. A. and G. R. Fauth. 1980. "Downtown Auto Restraint Policies: The Costs and Benefits for Boston." *Journal of Transport Economics and Policy* 14, no. 2, pp. 133–53.
- Goodwin, Phillip. 1987. "Long-term effects of public transport subsidy." In *Transport Subsidy*. Edited by S. Glaister. Bristol, U.K.: Policy Journals, pp. 121–27.
- Goss, R. O. and C. Jones. 1977. "The Economies of Size in Dry Bulk Carriers." In *Advances in Maritime Economics*. Edited by R. O. Goss. Cambridge, U.K.: Cambridge University Press.

Government of Canada. *Public Accounts of Canada*. Various issues.

Griliches, Z. 1972. "Cost Allocation in Railroad Regulation." *Bell Journal of Economics and Management Science* 3, pp. 26–41.

Hamilton, Gordon B. 1991. "Cost Competitiveness of Canadian Airports," paper presented to Air Transport Association of Canada 57th Annual General Meeting, Vancouver, Canada, November 10–12.

Haritos, Z. 1973. *Rational Road Pricing Policies in Canada*. Ottawa: Canadian Transport Commission.

— 1975. "Transportation Infrastructure Costs and Revenues in Canada." *Journal of Transport Economics and Policy*, January, pp. 16–33.

Harker, P. T. 1985. "Research Directions in Transportation Regulation and Pricing." *Transportation Research* 19A, no. 5, pp. 489–91.

Harmatuck, D. J. 1979. "A Policy Sensitive Railway Cost Function." *Logistics and Transportation Review* 15 (May), pp. 277–315.

Harris, R. 1977. "Economics of Traffic Density in the Rail Freight Industry." *Bell Journal of Economics and Management Science* 8, pp. 556–64.

Harrison B. 1986. "Electronic Road Pricing in Hong Kong." *Traffic Engineering and Control* 27, no. 1, pp. 13–18.

Hau, Timothy D. 1990. "Developments in Transport Policy: Electronic Road Pricing Development in Hong Kong 1983–1989." *Journal of Transport Economics and Policy*, May, pp. 203–14.

— 1991. *An Economic Analysis of Road Pricing: A Diagrammatic Approach*. Washington D.C.: World Bank. Transportation Division, Infrastructure and Urban Development Department, Policy Research and External Affairs.

Heaver, Trevor. 1970. "Tanker Freight Rates in the Short and Long Run." *Journal of Applied Econometrics* 10, pp. 203–17.

- Heaver, Trevor D. and T. Keast. 1974. "Transport subsidies in Canada." In *Transportation Subsidies — Nature and Extent*. Edited by K.M. Ruppenthal. Vancouver, B.C.: University of British Columbia, Centre for Transportation Studies, pp. 21–30.
- Heggie, I. G. 1989. *Public Sector Transport Pricing and Allocative Efficiency in Developing Countries*. Washington, D. C.: World Bank. Policy Planning and Research Staff Discussion Paper.
- 1991. *User Charging and Accountability for Roads: An Agenda for Reform*. Washington D. C.: World Bank Infrastructure Division, Discussion Paper.
- Hensher, David A. 1991. "Electronic Toll Collection." *Transportation Research* 25A, no. 1, pp. 9–16.
- Higgins, T. 1979. "Road Pricing — Should and Might it Happen?" *Transportation* 8, pp. 99–113.
- 1986. "Road-Pricing Attempts in the United States." *Transportation Research* 20A, pp. 145–50.
- Hoel, Lester, ed. 1990. "Financing Transportation Infrastructure." Special Issue, *Transportation Research* 24A, no. 4, July.
- Holland E. P. and P. L. Watson. 1978. "Traffic Restraint in Singapore." *Traffic Engineering and Control* 19, no. 1, p. 15.
- Hurley, W. 1989. *The Efficient Allocation of Runway Capacity*. Final Report. Transportation Development Centre Visiting Experts Program.
- International Road Federation. 1990. *World Road Statistics, 1985–1989*. Geneva.
- 1991. *World Road Statistics, 1986–1990*. Geneva.
- Interstate Commission. 1986. *An Investigation of Cost Recovery Arrangements For Interstate Land Transport*. Canberra: Australia General Public Service, April 1986.

- 1990. *Road Use Charges and Vehicle Registration: A National Scheme*, vol. 1. Canberra: Australia General Public Service, March 1990.
- Jordan, W. J. 1983a. "Heterogeneous Users and the Peak-Load Pricing Model." *Quarterly Journal of Economics* 93, pp. 127–38.
- 1983b. "The Theory of Optimal Highway Pricing and Investment." *Southern Economic Journal* 50, pp. 560–64.
- Jorgenson, Dale. 1992. "Fragile Statistical Foundations: The Macroeconomics of Public Infrastructure Investment." Mimeo. Harvard University, Department of Economics.
- Jorgenson, D. and Kun-Young Yun. 1990. "The Excess Burden of Taxation in the U.S." Discussion Paper No. 1528. Cambridge, Mass.: Harvard Institute of Economic Research.
- Kahn, A. 1970. *The Economics of Regulation*, vol. 1. New York: John Wiley & Sons.
- Keeler, T. 1972. "Airline Regulation and Market Performance." *Bell Journal of Economics and Management Science* 3, pp. 399–424.
- 1974. "Railroad Costs: Returns to Scale and Excess Capacity." *Review of Economics and Statistics* 56, pp. 201–08.
- Keeler, T. E., K. Small and Associates. 1975. *The Full Costs of Urban Transport, Part III: Automobile Costs and Final Intermodal Cost Comparisons*, Monograph No. 212. Berkeley: University of California, Institute of Urban and Regional Development.
- Keeler, T. E. and K. Small. 1977. "Optimal Peak-Load Pricing, Investment and Service Levels on Urban Expressways." *Journal of Political Economy* 85, pp. 1–25.
- Keeler, T. E., ed. 1985. *Research in Transportation Economics*, vol. 2. Greenwich, Connecticut: JAI Press Inc.

- Knight, F. H. 1924. "Some Fallacies in the Interpretation of Social Cost." *Quarterly Journal of Economics* 38, pp. 582–606.
- Koenker, R. 1977. "Optimal Scale and the Size Distribution of American Trucking Firms." *Journal of Transportation and Economic Policy* 11, pp. 54–67.
- Kraus, Marvin. 1981. "Indivisibilities, Economies of Scale, and Optimal Subsidy Policy for Freeways." *Land Economics* 57, no. 1, February, pp. 115–21.
- 1982. "Highway Pricing and Capacity Choice Under Uncertain Demand." *Journal of Urban Economics* 12, pp. 122–28.
- 1989. "The Welfare Gains from Pricing Road Congestion Using Automatic Vehicle Identification and On-Vehicle Meters." *Journal of Urban Economics* 25, pp. 261–81.
- Kraus, M., H. Mohring and T. Pinfeld. 1976. "The Welfare Costs of Nonoptimal Pricing and Investment Policies for Freeway Transportation." *American Economic Review* 66, pp. 532–47.
- Kulash, D. 1974. "Income Redistributational Consequences of Roadway Pricing." *Transportation Research Forum Proceedings* 15, pp. 153–59.
- Larsen O. I. 1988. "The Toll Ring in Bergen, Norway — The First Year of Operation." *Traffic Engineering and Control* 29, no. 4, pp. 216–22.
- Lave L. B. and J. S. DeSalvo. 1968. "Congestion, Tolls, and the Economic Capacity of a Waterway." *Journal of Political Economy* 76, pp. 375–91.
- Lee, Douglass B. 1982. "Net Benefits from Efficient Highway User Charges." *Transportation Research Record* 858, pp. 14–20.
- Levin, R. 1978. "Allocation in Surface Transportation: Does Rate Regulation Matter?" *Bell Journal of Economics and Management Science* 9, pp. 18–45.

- 1981. "Railroad Rates: Profitability and Welfare Under Regulation." *Bell Journal of Economics and Management Science* 11, pp. 1–26.
- Levine, M. 1969. "Landing Fees and the Airport Congestion Problem." *Journal of Law and Economics* 12.
- Lewis, W. A. 1949. *Overhead Costs: Some Essays in Economic Analysis*. New York: Rinehart.
- Likens, J. D. 1976. "The Welfare Costs of Nonoptimal Airport Utilization." *Journal of Public Economics* 5, pp. 81–102
- Little, I. M. D. and K. M. McLeod. 1974. "The New Pricing Policy of the BAA." In *Airport Economic Planning*. Edited by G. P. Howard, Cambridge, Mass.: The Massachusetts Institute of Technology, pp. 445–66. Reprinted from (1972), *Journal of Transport Economics and Policy* 6, pp. 101–15.
- Luck, D. P. and I. J. Martin. 1988. "Review of Road Cost Recovery." Occasional Paper 90. Canberra: Australia General Public Service, Bureau of Transport and Communications Economics.
- May, A. D. 1979. "Comment: Road Pricing — Should and Might It Happen?" *Transportation* 8, pp. 119–23.
- McShan, S. and R. Windle. 1989. "The Implications of Hub-and-Spoke Routing for Airline Costs and Competitiveness." *Logistics and Transportation Review* 35, no. 3 (September), pp. 209–30.
- Miyao, T. and P. Shapiro. 1981. "Discrete Choice and Variable Returns to Scale." *International Economic Review* 22, pp. 257–73.
- Mohring, Herbert D. 1970. "The Peak Load Problem with Increasing Returns and Pricing Constraints." *American Economic Review* 60, pp. 693–705.
- 1976. *Transportation Economics*. Cambridge, Mass.: Ballinger Publishing Company.
- 1984. "Profit Maximization, Cost Minimization, and Pricing for Congestion-Prone Facilities." *Logistics and Transportation Review*, pp. 27–36.

- Mohring, Herbert D. and I. Harwitz. 1962. *Highway Benefits: An Analytical Framework*. Evanston, Illinois: Northwestern University Press.
- Morrison, S. 1982. "The Structure of Landing Fees at Uncongested Airports: An Application of Ramsey Pricing." *Journal of Transport Economics and Policy* 16, pp. 151–61.
- 1983. "Estimating the Long Run Prices and Investment Levels for Airport Runways." In *Research in Transportation Economics*, 1. Edited by T. Keeler. Greenwich, Connecticut: JAI Press, pp. 103–31.
- 1984. "Equity, Efficiency and Individual Preference in the Pricing of Transportation Infrastructure." Mimeo. Evanston, Ill.: Northwestern University, Department of Economics.
- 1986. "A Survey of Road Pricing." *Transportation Research* 20A, pp. 89–97.
- 1987a. "The Theory of Optimal Highway Pricing and Investment: A Comment." *Southern Economic Journal* 53, no. 3, pp. 779–82.
- 1987b. "The Equity and Efficiency of Runway Pricing." *Journal of Public Economics* 34, pp. 45–60.
- Morrison, S. and C. Winston. 1986. *The Economic Effects of Airline Deregulation*. Washington, D.C.: The Brookings Institution.
- 1989a. "Airline Deregulation and Public Policy." *Science* 245, August, pp. 707–11.
- 1989b. "Enhancing the Performance of the Deregulated Air Transportation System." In *Brookings Papers on Economic Activity: Microeconomics*. Edited by C. Winston. Washington, D.C.: The Brookings Institution, pp. 61–125.
- 1990. "The Dynamics of Airline Pricing and Competition." *American Economic Review* 80, May, pp. 389–93.
- Nash, C. 1991. "Rail Policy in the Europe Community." Working paper 324. University of Leeds, Institute for Transport Studies, January.

- Nelson, James R. 1964. "Pricing Transportation Services." In *Transportation Investment and Economic Development*. Edited by Gary Fromm. Washington, D.C.: The Brookings Institution.
- Neutze, G. M. 1966. "Investment Criteria and Road Pricing." *Manchester School of Economics and Social Studies* 34.
- Newbery, D. 1988a. "Road User Charges in Great Britain." *Economic Journal* 98, pp. 161–76.
- 1988b. "Road Damage Externalities and Road User Charges." *Econometrica* 56, no. 2. pp. 295–316.
- Newell, G. 1987. "The Morning Commute for Non-identical Travellers." *Transportation Science* 21, pp. 74–88.
- Ng, Y. K. and W. Weisser. 1974. "Optimal Pricing with a Budget Constraint: The Case of the Two-Part Tariff." *Review of Economic Studies* 41, pp. 337–45.
- Nix, F. P. 1989. *Road-User Costs*, Report for Transport Canada, Economic Research Branch. Ottawa: Transport Canada.
- Nix, Fred P., Michel Boucher and Bruce Hutchinson. 1992. "Road Costs." In *Directions: Final Report of the Royal Commission on National Passenger Transportation*, vol. 4. Ottawa: Supply and Services Canada.
- Oum, Tae Hoon. 1981. "Efficiency Losses from the Unbalanced Recovery of Costs of Transportation Infrastructure in Canada." *Logistics and Transportation Review* 18, no. 1.
- Oum, Tae Hoon and Chunyan Yu. 1990. "An International Comparison of the Economic Efficiency of Passenger Railway Systems." A report prepared for the Royal Commission on National Passenger Transportation, RR-08, December.
- Oum, Tae H. and Michael W. Tretheway. 1988. "Ramsey Pricing in the Presence of externality Costs." *Journal of Transport Economics and Policy* 22, no. 3, pp. 307–17.

- Oum, Tae H. and Yimin Zhang. 1990 "Airport Pricing: Congestion Tolls, Lumpy Investment and Cost Recovery." *Journal of Public Economics* 43, pp. 353–74.
- Panzar, J. and R. Willig. 1977. "Free Entry and the Subsistability of Natural Monopoly." *Bell Journal of Economics and Management Science* 8, pp. 1–22.
- Park, R. E. 1989. *Incremental Costs and Efficient Prices with Lumpy Capacity*. Santa Monica, Calif.: Rand Corporation.
- Phillips, A. 1980. "Ramsey Pricing and Sustainability with Interdependent Demands." In *Regulated Industries and Public Enterprise*. Edited by B. M. Mitchell and P. R. Kleindorfer. Lexington, Mass.: D.C. Heath.
- Pigou, Arthur C. 1912. *Wealth and Welfare*. London: Macmillan Press.
- 1920. *The Economics of Welfare*. London: Macmillan Press.
- Rees, Ray. 1976. *Public Enterprise Economics*. London: Weidenfeld and Nicolson.
- Roth, G. J. 1965. "Economic Approach to Road Congestion." *Town Planning Review* 36, pp. 49–61.
- Roy, R. 1980. *Economics of Scale in the Airline Industry*. Discussion paper. Ottawa: Canadian Transport Commission.
- Royal Commission on National Passenger Transportation. 1991. *Getting There: The Interim Report of The Royal Commission on National Passenger Transportation*. Ottawa: Supply and Services Canada.
- Ruggles, Nancy. 1949-50. "The Welfare Basis of the Marginal Cost Pricing Principle." *Review of Economic Studies* 17, no. 1.
- 1949-50. "Recent Developments in the Theory of Marginal Cost Pricing." *Review of Economic Studies* 17, no. 2.

- Sarndal, C., T. Oum and W. Statton. 1978. "Further Evidence on Factors Influencing Operating Costs of U.S. Commercial Airlines." *Journal of Transport Economics and Policy* 12, pp. 47–55.
- Shah, Anwar M. 1990. "Optimal Pricing of Traffic Externalities; Theory and Measurement." *International Journal of Transport Economics* 17, pp. 3–19.
- Sharkey, W. W. 1982. "Suggestions for Game-theoretic Approach to Public Utility Pricing and Cost Allocation." *Bell Journal of Economics and Management Science* 13, pp. 57–68.
- 1985. "Economic and Game Theoretic Issues Associated with Cost Allocation in Telecommunications Networks." In *Cost Allocation: Methods, Principles and Applications*. Edited by P. H. Young. Amsterdam: Elsevier Science Publishers.
- Small, K. 1983a. "The Incidence of Congestion Tolls on Urban Highways." *Journal of Urban Economics* 13, pp. 90–111.
- 1983b. "Bus Priority and Congestion Pricing on Urban Expressways." *Research in Transportation Economics* 1, pp. 27–74.
- 1990a. "Highway Infrastructure: Crisis in Finance or Crisis in Management?" *Institute of Transport Studies Review* 13, no. 2 (February), pp. 1–3.
- 1990b. Urban Transport Economics. Draft monograph.
- Small, K. and C. Winston. 1986. "Efficient Pricing and Investment Solutions to Highway Infrastructure Needs." *American Economic Review* 76, no. 2 (May), pp. 165–69.
- 1988. "Optimal Highway Durability." *American Economic Review* 78, no. 3, pp. 561–69.
- Small, K., C. Winston, and C. Evans. 1989. *Road Work: A New Highway Pricing and Investment Policy*. Washington, D.C.: The Brookings Institution.

- Spady, R. and A. Friedlaender. 1978. "Hedonic Cost Functions for the Regulated Trucking Industry." *Bell Journal of Economics* 4, pp. 159–79.
- Spulber, D. 1989. *Regulation and Markets*. Cambridge, Mass.: The MIT Press.
- Starkie, D. 1982a. "Road Indivisibilities: Some Observations." *Journal of Transport Economics and Policy*, September, pp. 259–66.
- 1982b. "Cost Recovery: An Investment Perspective." In *Pricing and Cost Recovery in Long Distance Transport*. Edited by D. N. M. Starkie et al. The Hague: Martinus Nijhoff.
- Starrs, M. M. and David Starkie. 1986. "An Integrated Road Pricing and Investment Model: A South Australian Application." *Australian Road Research* 16, pp. 1–9.
- Statistics Canada. *Consolidated Government Finance*, Catalogue No. 68-202, Table 1.
- Steiner, P. O. 1967. "Peak Loads and Efficient Pricing." *Quarterly Journal of Economics* 71, pp. 585–610.
- Strotz, R. 1965. "Urban Transportation Parables." In *The Public Economy of Urban Communities*. Edited by Julius Margolis. Baltimore, MD.: Johns Hopkins Press, pp. 127–69.
- Stubbs, P. W., W. J. Tyson and M. Q. Dalvi. 1980. *Transport Economics*. London: George Allen and Urwin Publications.
- Toms, M. 1986. "Towards Cost Related Pricing." Polytechnic of Central London, School of the Environment. Mimeo.
- Train, K. 1977. "Optimal Transit Prices Under Increasing Returns to Scale and a Loss Constraint." *Journal of Transport Economics and Policy* 11, no. 1, pp. 185–94.
- 1991. In *Pursuit of Optimal Regulation: The Economic Theory of Natural Monopoly*. Cambridge, Mass.: The MIT Press.

- Transport Canada. 1990. *Proposed New Cost Recovery Policy*, Phase II. Discussion paper. No. TP 10041, Ottawa, April.
- 1991. *Air Services Charges Regulations*, Cost Recovery Branch, Finance and Administration, Fifth Edition. No. TP-2590E, Ottawa, April.
- Turvey, R. 1968. "Peak-Load Pricing." *Journal of Political Economy* 76, February, pp. 101–13.
- U.S. Congressional Budget Office. 1992. *Paying for Highways, Airways, and Waterways: How Can Users be Charged?* Washington, D.C.
- U.S. Department of Commerce, Bureau of the Census. 1989. *The Statistical Abstract of the United States*. T.1018, The National Data Bank, Transport section. Washington, D.C.
- U.S. Department of Transportation. 1990. *Moving America: New Directions, New Opportunities*. Washington, D.C., February.
- 1991. "Air Services Charges Regulations," 5th ed., April 1.
- Vickery, W. 1965. "Pricing as a Tool in the Coordination of Local Transportation." In *Transportation Economics*. New York: Columbia University Press, National Bureau of Economic Research.
- 1968. "Congestion Charges and Transportation Investment." *American Economic Review*, Papers and Proceedings, 59, pp. 107–18.
- 1985. "The Fallacy of Using Long-run Cost for Peak-Load Pricing." *Quarterly Journal of Economics* 100, pp. 1331–34.
- Viton, P. A. 1980. "Equilibrium Short-run Marginal Cost Pricing of a Transport Facility: The Case of the San Francisco Bay Bridge." *Journal of Transport Economics and Policy* 14, pp. 185–203.
- Walters, Alan A. 1961. "The Theory and Measurement of Private and Social Cost of Highway Congestion." *Econometrica* 29, no. 4 (October), pp. 676–99.

- 1968. *The Economics of Road User Charges*. World Bank Occasional Papers No. 5. International Bank for Reconstruction and Development. Baltimore, MD.: Johns Hopkins University Press.
- 1983. "Airports: An Economic Survey." *Journal of Transport Economics and Policy* 12, pp. 125–60.
- Westfield, F. 1966. "Practicing Marginal Cost Pricing — A Review." *Journal of Business* 39, no. 1, part 1, January, pp. 57–74.
- Wilson, George W. 1962. *Essays on Some Unsettled Questions in the Economics of Transportation*. Bloomington: Indiana University Press.
- 1980. *Economic Analysis of Intercity Freight Transportation*. Bloomington: Indiana University Press.
- Winston, C. 1985. "Conceptual Developments in Economics of Transportation: An Interpretive Survey." *Journal of Economic Literature* 23, pp. 57–94.
- 1991. "Efficient Transportation Infrastructure Policy." *Journal of Economic Perspectives* 5, no. 1, Winter, pp. 113–27.
- Wohl, M. and C. Hendrickson. 1984. *Transportation Investment and Pricing Principles*. Toronto: J. Wiley & Sons.
- Ying, J. S. 1990. "The Inefficiency of Regulating a Competitive Industry: Productivity Gains in Trucking Following Reform." *The Review of Economics and Statistics* 72, no. 2, (May), pp. 191–201.
- Young, Peyton H., ed. 1985. *Cost Allocation: Methods, Principles, Applications*. Amsterdam: Elsevier Science Publishers.
- Zajac, E. E. 1974. "Note on an Extension of the Ramsey Inverse Elasticity of Demand Pricing or Taxation Formula." *Journal of Public Economics* 3, pp. 181–84.

COMPETITION POLICY AND CANADIAN PASSENGER TRANSPORTATION

John Blakney*

October 1991

I. INTRODUCTION

This report examines the effectiveness of the *Competition Act* as an instrument to protect the interests of consumers of commercial air, rail and bus passenger transportation services relative to:

- (1) existing direct regulatory controls established by the *National Transportation Act, 1987* (NTA), the *Railway Act* (RA) and the *Motor Vehicle Transport Act, 1987* (MVTA); and
- (2) generally, the more interventionist approaches of public utility regulation of prices, profits and products where prior approvals of a government body for some or all of these business activities are required by legislation. State ownership of privately owned commercial passenger transportation services is not examined directly as an alternative regulatory mechanism.

This is a study of the choice of governing instrument where the focus is government intervention to restrict private business choice in situations where unconstrained business choice would not be in the public interest.

* Fraser & Beatty.

Given that the public interest is multi-faceted and mercurial in a pluralistic democracy, this evaluation must look both at statements of purpose in enabling legislation and the ongoing application of the legislation's standards.

Accordingly this report examines the interrelationship between:

- (1) the substantive commercial conduct standards and remedies of the *Competition Act* and other more direct regulatory mechanisms with respect to the passenger transportation sector; and
- (2) the institutional and administrative structures for applying these substantive standards and remedies.

The particular aspect of the public interest selected as the focus for evaluation in this study is the "consumer interest." There is, unfortunately, no single pervasively accepted set of standards defining what business conduct is in the consumer interest and what business conduct is not. "Consumer interest" has become a political term to a large extent. The definition of consumer interest selected for this study must therefore be somewhat arbitrary and it certainly will not please everyone.

This study assumes that the consumer interest comprehends maximizing product choice and innovation and minimizing opportunities for businesses to be less than fully efficient in the supply of their products.

In this context, achieving fair prices through internal cross subsidization is regarded only as a means to achieve maximum consumer choice and business efficiency and not as an end in itself.

Inclusion of social goals or income distribution, policy-pricing objectives as a necessary and important aspect of the consumer interest would, therefore, result in considerably different analysis and results. As discussed below, in parts II, III and IV, fair or politically motivated pricing is not an objective of the *Competition Act*. However, existing passenger service regulation creates both the incentive and the opportunity to establish prices based predominantly on fairness and social policy either directly or through controlling exit and entry.

It should, however, be noted that unregulated markets are rife with prices that are established not just to cover marginal costs. There are other pricing considerations related to sales maximization, stimulating demand for a firm's complementary or related products, administrative convenience, and maintenance of good customer relations. From the economist's strict marginal cost pricing perspective, all these can contribute to unregulated supplier price structures which display stability and geographical uniformity and which entail an element of internal cross subsidization.

Private markets regulated only by the *Competition Act* can and do deliver price structures that are fair, stable and, to some degree, geographically uniform in consumers' eyes. Such markets do not, however, deliver price structures that involve substantial redistributions of wealth geographically between particular income groups.

To summarize, this report examines alternative regulatory instruments in relation to their capacity to facilitate maximum consumer choice and business efficiency in the supply of commercial passenger transportation services. It does not look at their capacity to function as taxation and subsidization instruments.

II. *COMPETITION ACT*

A. OVERVIEW

The *Competition Act* is general legislation of general application founded principally upon the federal trade and commerce power. Criminal prohibitions in the Act against agreements to lessen competition (the original focus of Canadian competition law) have also been supported by the federal criminal law power.

The legislation is enforced by the Director of Investigation and Research (DIR), a Governor in Council appointee responsible to the Minister of Consumer and Corporate Affairs. In practice, the DIR's office, known as the Bureau of Competition Policy (Bureau), operates with a measure of independence from political control similar to a combination of a provincial police force and Crown Attorney's office. The DIR is responsible for the conduct of all formal inquiries into possible criminal offences and practices reviewable

by the Competition Tribunal and for making applications for remedial orders to the Competition Tribunal. Criminal matters are generally referred to the Attorney General for Canada for the laying of charges and prosecution, but the Attorney General's Agent generally consults closely with the DIR on these matters. The DIR reports annually to Parliament through the Minister.

The legislation has been broadened, refined and clarified a number of times over its 100-year history, with the most recent and comprehensive revision in 1986. The 1986 revisions:

- (1) clarified the conspiracy prohibitions in relation to the case law and expanded opportunities for lawful export agreements;
- (2) established the Competition Tribunal (to replace the Restrictive Trade Practices Commission (RTPC)) as an expert quasi-judicial body comprising Federal Court judges and business and economics experts to adjudicate reviewable practices. (The Competition Tribunal is purely an adjudicative body and, unlike the RTPC, does not have the capacity to support DIR inquiries or conduct fact-finding inquiries into monopolistic practices);
- (3) revised the Act's investigative powers to comply with the *Canadian Charter of Rights and Freedoms*;
- (4) substituted new reviewable practices respecting mergers and abuse of dominant position for ineffective criminal law prohibitions against mergers and monopolies;
- (5) introduced new reviewable practices respecting systematic delivered pricing and specialization agreements;
- (6) established a new purposes section to guide administration of the Act by both the DIR and the Competition Tribunal; and
- (7) established a new requirement to notify the DIR of mergers above a certain threshold to assist in enforcement of the merger provisions.

As noted, the substantive standards of the *Competition Act* comprise both criminal law prohibitions administered by the provincial superior courts, and reviewable practices over which the Competition Tribunal has sole jurisdiction to issue remedial Orders.

The principal criminal prohibitions, that is, offences in relation to competition, relate to:

- agreements to lessen competition unduly in a market;
- bid-rigging;
- resale price maintenance;
- price discrimination among competing purchasers;
- predatory pricing; and
- misleading advertising.

Matters reviewable by the Competition Tribunal include:

- non-price vertical restraints (refusal to deal, consignment selling, tied selling, market restriction and exclusive dealing) engaged in by major participants;
- abuse of dominant position which substantially lessens competition in a market;
- systematic delivered pricing;
- specialization agreements; and
- mergers which substantially lessen competition in a market.

These categories are not entirely self-contained. It is possible to maintain an abuse of dominant position application using evidence of predatory pricing or non-price vertical restraints, and it is possible to support a merger application based upon increased prospects for post-merger cartelization or abuse of dominance.

The foundations of Canadian competition law, in my view, are the prevention of:

- (1) inefficient monopolization or increased market concentration whether through acquisition or horizontal agreement; and
- (2) exercising of market power to exclude competition in primary markets, in secondary markets or in new markets, whether through pricing or contractual practices.

These are also the basic objectives of competition law throughout the world.

Maximizing competition is viewed as a necessary and a generally sufficient condition to maximizing consumer welfare in the classical economics literature.

Notwithstanding the classical economics underpinnings of the *Competition Act*, the 1986 revisions established a more human and accessible face to the law through a new purposes provision (section 1.1):

The purpose of this Act is to maintain and encourage competition in Canada in order to promote the efficiency and adaptability of the Canadian economy, in order to expand opportunities for Canadian participation in world markets while at the same time recognizing the role of foreign competition in Canada, in order to ensure that small and medium-sized enterprises have an equitable opportunity to participate in the Canadian economy and in order to provide consumers with competitive prices and product choices.

Under the regulatory intervention model of competition law, markets are presumed to be functioning properly unless certain standards are breached. In such cases, the State is considered to be entitled to intervene to remedy a defined problem. The remedy can be justified in lasting only long enough to eliminate the problem and to restore adequate competition in the market. The remedy is imposed by the courts or an adjudicative body; it is not legislated.

As a result, the application of competition law remedies occurs in a strongly contested adjudicative environment where the person(s) against whom the remedy is sought have extensive rights to challenge the government position. In the event a remedy is authorized, they may return at their own instance to the adjudicative environment to seek modification or elimination of the remedy if it can be shown to be no longer effective or even counter-productive.

Debate on the relative merits of competition law and more interventionist direct regulatory instruments, therefore, turns not so much on the structural coverage and efficiency of competing instruments to remove consumer harms once the State acts, but on competing political or even ideological visions on the appropriate relationship of the State and private business.

Competition law, with its built-in presumptions that the onus is on the State to prove that focussed intervention is justified in the eyes of an impartial adjudicator on a case-by-case basis, strikes a political or ideological chord with those who contend that government intervention in business decision-making is inherently counter-productive and that social goals should be achieved through taxation, subsidization and the State supply of pure public goods, that is, products which society demands but which private markets fail to supply. Competition law also finds favour with those who believe that general legal standards can and will be followed by business without direct and specific State instruction, and that there is a tangible general deterrent effect of obtaining litigated remedies against participants in the economy.

Direct regulation, on the other hand, is supported by those who consider that private enterprise, because of pervasive market imperfections including imperfect information, imperfect price signals and a myriad of supply rigidities, can and should generally benefit from State guidance or even control of business management. Direct intervention is also favoured by those who consider that it is not feasible to construct general economic conduct laws that provide useful standards for action by particular businesses.

This political or ideological cleavage becomes particularly pronounced in an examination of the role of the State in relation to the supply of products which are important to all consumers: food, housing, transportation, communications and energy. The ideological gap is even more evident when the inputs required to produce these products include a high proportion of priced or unpriced products supplied by the State, that is, public goods.

This report, however, does not attempt to address or reconcile these political or ideological differences in examining the relative effectiveness of the *Competition Act*. Rather, the analysis flows from the author's perspective that competing governing instruments should be evaluated solely upon their *a priori*, or demonstrated, capacity to achieve certain objective performance standards.

B. SCOPE OF COVERAGE

The *Competition Act* is a general law of general application. The key terms of the Act (product, business, supply, trade, industry profession) are defined with maximum breadth and flexibility. However, its application to economic conduct in Canada is not universal. Statutory and common law rules limit its application.

With respect to passenger transportation services, the key limitations relate to both government and regulated private sector conduct.

1. Statutory Exemptions

Given the constitutional underpinnings of the *Competition Act*, it is generally accepted that only validly enacted federal legislation may expressly exempt an activity from the operation of the *Competition Act*. The instances in which Parliament has elected to limit the scope of the *Competition Act* are few and far between.

Agreements among shipping conference members are exempted from the conspiracy prohibition of the *Competition Act* if certain conditions are met by operation of the *Shipping Conferences Exemption Act, 1987*. The *Competition Act* itself contains certain limited exemptions relating to activities that need not be considered for the purpose of this report. The *Farm Product Marketing Agencies Act* also contains an express exemption with respect to the conspiracy prohibition.

2. Common Law Exemptions

(a) Crown Agency

It is a general rule of statutory interpretation applied by the courts that an enactment is not binding on the Crown or its agents unless the enactment expressly says that it is. Crown agency status may be expressly granted by valid legislation of the federal government or a province, or it may be a constructive agency arising from the facts — the actual relationship between a person and the Crown.

This principle has been applied by the courts to the *Combines Investigation Act* (the predecessor legislation to the *Competition Act*) with respect to

prosecutions under the Act's criminal conspiracy provisions (*R. v. Eldorado Nuclear Ltd.*, [1983] 2 S.C.R. 551, 4 D.L.R. (4th) 193). It is also established that this principle applies equally to federal legislation of a non-criminal and regulatory nature founded upon the federal trade and commerce or interprovincial undertakings power, for example, *Alberta Government Telephone v. CRTC* (1989), 61 D.L.R. (4th) 193 (SCC).

The *Competition Act* expressly applies to Crown Agents that are corporations *but only* to the limited extent set out in section 2.1 of the Act:

This Act is binding on and applies to an agent of Her Majesty in right of Canada or a province that is a corporation, in respect of commercial activities engaged in by the corporation in competition, whether actual or potential, with other persons to the extent that it would apply if the agent were not an agent of Her Majesty.

Therefore the *Competition Act* does not apply to, and would not be a substitute for, valid regulatory legislation affecting:

- (i) the federal and provincial Crowns themselves acting other than through a corporation; and
- (ii) federal or provincial Crown corporations which are either not engaged in commercial activities or are not engaged in activities which are in actual or potential competition with other persons in Canada.

Stated another way, without new federal legislation expressly making the *Competition Act* applicable, the *Competition Act* currently does not apply to:

- (i) commercial activities conducted by federal or provincial government departments;
- (ii) activities of Crown corporations that are not regarded by the courts or the Competition Tribunal (depending upon the activity in question) as being "commercial activities." (There is no definition of commercial activity in the *Competition Act*, and the Bureau of Competition Policy has provided no administrative guidance to date with respect to its interpretation of this term); and
- (iii) commercial activities of a Crown corporation which are not engaged in competition with others. (Again, there is no statutory or judicial

amplification of what constitutes a non-competitive commercial activity, but presumably this might include the monopoly supply of products in a given market. From the wording of section 2.1, it is possible to argue that only actual and not potential competition is relevant.)

These restrictions clearly have implications for the effectiveness of the *Competition Act* where governments consider the privatization, devolution, contracting out or pricing of elements of the transportation system which have heretofore been supplied by government departments as pure or priced public goods. Such elements include airport landing slot or runway access, road access or use, provision of navigational services and the provision of regulatory inspection services. Accordingly, we will examine the concept of Crown agency more fully in Part V of this report.

(b) Regulated Conduct

It is a well-established principle that business conduct that is required or authorized by validly enacted federal or provincial legislation is not subject to the *Competition Act* (unless the legislation also expressly makes the *Competition Act* applicable). There is some debate as to whether this doctrine is, in effect, a constitutional doctrine based upon paramountcy considerations rendering application of the *Competition Act* invalid in principle; or whether it amounts instead to a regulated conduct defence whereby no sanction or remedy will be issued if the accused (or the respondent as the case may be) demonstrates as a matter of fact that an element the federal government must prove cannot in the circumstances be proven because the accused relied upon prior government requirements or approvals.

This is by no means a purely academic legal distinction. If the regulated conduct doctrine is quasi-constitutional, then both the inquiry and remedial powers of the *Competition Act* do not apply to a particular industry or activity, and efforts to amend the Act to expand its scope, and thereby to trench upon provincial or federal activities from the trade and commerce power base of the legislation, would be ineffective. Equally the provinces could essentially enact any regulatory legislation that would otherwise be valid under provincial heads of power to ensure that the *Competition Act* could not apply to the subject activity. Probably such provincial action could extend only to intra-provincial activities. The leading case in support of this constitutional perspective is *A.-G. Canada v. LSBC*, [1982] 2 S.C.R. 307, 137 D.L.R. (3d) 1.

If there is only a regulated conduct defence, then the inquiry powers of the Act would still be valid and the courts or the Competition Tribunal, as the case may be, would have greater scope, first, to ascertain whether the regulatory scheme was effective in requiring the impugned behaviour, and, secondly, where the behaviour was merely authorized, to weigh the relative merits of allowing or not allowing the defence in relation to the goals of the competing legislation. The leading case favouring this approach is *R. v. Canadian Breweries Ltd.* (1960), 126 C.C.C. 133, [1960] O.R. 601, 33 C.R. 1 (H.C.J.).

In this context both the LSBC and the Canadian Breweries cases determined the applicability of criminal prohibitions of the former *Combines Investigation Act*. There have been no successful constitutional challenges to date regarding the reviewable practices provisions of the *Competition Act* which are exclusively founded upon the federal trade and commerce power. As well, recently, the Supreme Court has been more open-minded in determining the scope of the trade and commerce power. In *A.-G. Canada v. C.N. Transportation Ltd.*, [1983] 2 S.C.R. 206, 3 D.L.R. (4th) 16, the Court upheld the validity of the non-price vertical restriction practices of the Act and ruled that they were not an excessive invasion of the provincial property and civil rights domain.

Nevertheless, the regulated conduct doctrine is the basis for defining the applicability of the *Competition Act* in relation to the regulation of commercial passenger service entry, exit, merger and price. In my view, the current state of the law in relation to this boundary is as follows:

- (i) It is now uncertain whether licensing requirements for *entry or exit or ownership* based on economic (that is, public interest, or public convenience and necessity) considerations as opposed to commercial fitness criteria (that is, fit, willing and able) could by themselves make the merger or abuse of dominant position provisions of the Act inapplicable to the subject industry. However, licensing and merger approval involving detailed conditions or performance requirements which clearly constrain business behaviour would very likely make the merger and abuse of dominance provisions inapplicable at least with respect to a case based upon a lessening of competition in the activities caught by the conditions or performance requirements.

- (ii) *Price approval requirements or a power of price disallowance* coupled with statutory pricing standards in regulatory legislation would make the predatory pricing and price discrimination provisions of the *Competition Act* inapplicable to unilateral pricing activities of regulated companies. However, low pricing, even if sanctioned by a regulator, could still constitute anti-competitive practice for the purpose of establishing abuse of dominant position.
- (iii) The *conspiracy prohibitions* of the Act may be inapplicable only where the impugned agreement or arrangement is specifically required or authorized as part of an otherwise valid regulatory scheme, or the agreement or arrangement is necessary to the achievement of valid regulatory purposes.
- (iv) For substantive provisions of the Act apart from conspiracy, abuse of dominant position, predatory pricing and merger, there is no regulated conduct defence or exemption for regulated industries.

C. ENFORCEMENT POWERS AND PRACTICES

Inquiries by the DIR may be either informal or formal, with the boundary line being somewhat unclear.

An informal inquiry technically becomes a formal inquiry where:

- (1) the DIR has received a six-resident complaint in proper form as required by section 9 of the Act;
- (2) the DIR has reason to believe that grounds for a remedial order exist or an offence under the Act has been or is about to be committed; or
- (3) a formal inquiry is directed by the Minister.

The Act requires all formal inquiries to be private and, in practice, all informal inquiries are conducted also in private.

Most inquiries (with the possible exception of merger inquiries) are based upon complaints from competitors, suppliers or purchasers. Because of the Act's privacy requirement, complainants are not kept abreast of the inquiry they have initiated and may not be involved (except with the consent of the subject of the inquiry) in negotiations directed at an out-of-court settlement.

The Act provides special, superior court-supervised information-gathering powers to support formal inquiries which are subject to judicial supervision. These include powers to search and seize and to require oral examination, production and written returns.

The inquiry process is generally time consuming. There are no statutory deadlines that must be observed. Some major conspiracy inquiries have taken several years before charges were laid. The inquiry stages of the three predatory pricing cases that have gone to trial under the criminal provisions of the Act took many months, nor was interim prohibition sought to stop the alleged predatory conduct, as is provided for in the Act. Recently, reviewable practices inquiries have taken several months to a year before applications were filed with the Competition Tribunal.

Merger inquiries are, however, fast-tracked to ensure that the parties know whether the DIR intends to challenge the merger before its implementation.

The adjudicative process has also proven to be fairly time consuming. The determination of criminal charges (apart from those arising from misleading advertising and resale price maintenance) has often taken from one to several years. This is the case in part because of the criminal law requirement for a preliminary inquiry and a high incidence of procedural and Charter challenges raised by those accused. Recent Competition Tribunal proceedings have ranged from 12 to 18 months from application to decision.

Because of the costs and uncertainties of the inquiry and adjudicative process, the DIR in particular, but also the parties to the inquiry, have strong incentives to reach out-of-court settlements. Settled court-approved prohibition orders with respect to criminal conspiracy inquiries and informal merger undertakings have been increasingly prevalent in the last five years. Less frequently, the DIR has sought Competition Tribunal approval of a merger agreement (*DIR v. Palm Dairies Ltd.* (1986), 12 C.P.R. (3d) 540 (Comp. Trib.) and *DIR v. Imperial Oil Ltd.* (unreported Comp. Trib. decision, February 6, 1990), CT-89/3). One disputed merger, *DIR v. Air Canada* (1989), 27 C.P.R. (3d) 476 (Comp. Trib), resulted in a consent agreement following the application to the Tribunal but before trial.

Complainants have limited opportunity to participate in trials under the Act. In criminal matters complainants may only participate as witnesses, usually for the Crown. In reviewable matters, the Competition Tribunal has the

discretion to allow third-party interventions (*American Airlines v. Canada* (Comp. Trib.), [1989] 2 F.C. 88, 23 C.P.R. (3d) 178, 54 D.L.R. (4th) 741 (C.A.), affd [1989] 1 S.C.R. 236) but has been fairly tough in attaching conditions to intervener participation. These conditions include having to ask the DIR to call certain evidence before being allowed to do so, and avoidance of repetitive cross-examination. On the other hand, intervener participation in hearings to approve the Imperial/Texaco merger conditions and the Reservec merger case was thorough and extensive. These competitor interventions had considerable impact on the outcome.

The most severe criticisms of the *Competition Act* are directed, not surprisingly, at the inquiry and adjudicative processes of the regime, and not at the Act's substantive standards. They relate to the secretiveness of the inquiry process, the very long time from complaint to result, the low incidence of complaint-induced results, and the lack of complainant opportunity to participate in inquiries or in adjudication.

In short, justice is not seen to be done, and justice delayed is justice denied (especially when the acts in question have affected the market before the DIR appears to do anything). Criticism is also focussed on the inadequacy of the government-led anti-trust machinery because, unlike the U.S., there is no tradition (and little apparent opportunity or incentive) for private anti-trust enforcement. In the U.S., private anti-trust suits by far outweigh government suits — a reflection to some degree of the availability of treble damages and contingency fees.

However, within a broad range, it is possible to reform the inquiry and adjudicative process without undermining or affecting the substance of the law. It might also be possible to restructure a large part of the criminal law substance of the Act as civilly reviewable practices given the clear overall current foundation of the law on the trade and commerce power.

III. *COMPETITION ACT* — PRINCIPAL JURISPRUDENCE

A. INTRODUCTION

In light of the inquiry and adjudicative process, it should not be surprising that there is a very low volume of case law relating to the key provisions

concerning monopolistic and cartel behaviour. There is also a dearth of case law relating specifically to the transportation sector.

This section examines the core provisions of the Act that interface with direct regulatory schemes. They are:

- predatory pricing;
- conspiracy;
- merger; and
- abuse of dominant position.

The analysis covers recent judicial and Competition Tribunal decisions. In the case of merger and predatory pricing, recently published proposed, DIR enforcement guidelines are discussed in detail. With respect to conspiracy, the focus is more on the substance of recent prohibition orders agreed to in lieu of trial. Two important recent prohibition orders relate to the used household goods transportation industry, and the for-hire general trucking industry in Western Canada.

The published DIR analysis of the clearance of the CAIL/Wardair merger sheds some light on merger review. It should be noted that the National Transportation Agency and the Competition Tribunal have concurrent jurisdiction over airline mergers.

In addition, the Reservec/Gemini merger proceeding addresses, through a transportation service, the application of competition law to the consequences of mergers establishing control over the supply of essential or “bottleneck” inputs consumed by them and their competitors.

On the other hand, there are no reported Canadian competition law decisions or settled cases relating specifically to the passenger transportation sector where the DIR has actually challenged industry conduct using the enforcement machinery of the Act. The DIR has, of course, been a strong proponent of economic deregulation of this sector (with some particular success in air transport policy) and has also advocated privatization and pricing of quasi-public goods as necessary steps to establish self-sustaining and efficient markets in the supply of these services.

B. PREDATORY PRICING AND PRICE DISCRIMINATION (PARAGRAPH 50(1)(C) OF THE *COMPETITION ACT*)

1. Introduction

Section 50 of the *Competition Act* establishes three separate but related criminal offences addressing unilateral pricing behaviour.

The Act's *predatory pricing* standard provides:

Everyone engaged in a business who . . .

- (c) engages in a policy of selling products at prices unreasonably low, having the effect or tendency of substantially lessening competition or eliminating a competitor, or designed to have that effect,

is guilty of an indictable offence and liable to imprisonment for a term not exceeding two years.

Section 50 also contains prohibitions against *price discrimination* in sales of “articles” to competing purchasers, and against *geographic price discrimination* which substantially lessens competition.

Although the price discrimination offence is restricted to sales or articles, article is defined in the Act to include tickets or like evidence of a right to transportation. The price discrimination offence therefore, in my view, applies to sales of transportation services. However, as noted, the discriminatory sales must be to competing purchasers. This requirement has been interpreted consistently by the Bureau of Competition Policy to mean purchasers which compete with each other in the product they produce, or purchasers which are major purchasers of an article in limited supply.

Neither test for competing purchasers would appear to apply realistically to the sale of passenger transportation services with the exception of sales or services to brokers, wholesalers or retail resellers (for example, tour operators, travel agents) who compete in the same product or geographic market.

Accordingly, this report does not focus on the price discrimination prohibition except to note that it has limited applicability to the passenger service intermediary sector. It should not, in its present form, be regarded as a

means of preventing price discrimination at the retail level of the direct sale of commercial passenger services.

In addition, section 50 as presently drafted does not apply to price discrimination in sales of transportation services to discrete geographic monopolies.

Nor, and this is the principal limitation of the section, does this prohibition apply to the sale of services which are inputs to, or products of, commercial passenger transportation service suppliers. For example, the price discrimination prohibition would not apply to the pricing of computerized reservation services or travel services such as car rentals or hotels packaged with commercial passenger transportation services.

The geographic price discrimination offence has not been applied on a stand-alone basis. Complaints of geographical price discrimination are examined by the DIR in the context of either predatory pricing or abuse of dominant position.

It is too early to tell whether the predatory pricing offence will stay separate for enforcement purposes, as it is possible to structure a predatory pricing case involving a dominant firm (the only kind of firm that economic theory suggests would have the incentive and ability to predate successfully) as an abuse of dominance case. This would mean that the criminal standard of proof would be avoided, and the hearing would be before an expert tribunal.

The history of predatory pricing charges (three trials, two not guilty findings, one conviction where the price was zero,¹ and an apparent incapacity of the courts to focus individually on the elements of the offence, thus providing no useful guidance to the Bureau or to producers) would suggest that the Bureau may well prefer to pursue predatory pricing issues before the Competition Tribunal in the future.

The history of predatory pricing charges and inquiries over the last 20 years also suggests that Canadian competition law authorities have been affected by the strong criticism levelled by senior U.S. economists against the presumption of the offence that firms (even dominant ones) have an incentive and an ability to increase long-term profits through predatory pricing, regardless of what production cost-based price floor for the determination

of “unreasonably low prices” may exist. As this critique goes, predatory pricing might result in a long-term loss of consumer welfare only where the industry is subject to significant long-term barriers to entry.

This uncertainty has, in my view, been reflected in the draft Predatory Pricing Bulletin discussed below.

The following discussion of the Bureau’s proposed predatory pricing enforcement policy can also be applied to Bureau enforcement of the abuse of dominant position provision where the focus is the unilateral pricing behaviour of the dominant firm.

In addition, and perhaps more important, those same criteria that are presented in the draft Predatory Pricing Bulletin to determine whether structural market conditions exist to exercise short-run market power could be used by the Bureau to assess the related issues of market dominance and substantial lessening of competition under the abuse of dominant position provisions.

2. Draft Predatory Pricing Bulletin

In April 1990, the Bureau circulated for public comment a draft Bulletin laying out its enforcement policy with respect to the predatory pricing offence of the *Competition Act*.²

The Bulletin notes at the outset that there is very limited jurisprudence on the interpretation of this provision and that, in the past, the Bureau has provided little public guidance on its enforcement policy regarding predatory pricing.

The Bulletin restyles the predatory pricing offence somewhat by defining it as the sale of products at prices so low as to cause injury to competition through the elimination of a competitor or the deterrence of entry or expansion of a competitor. Injury to competition is defined as a situation where the alleged predator is regarded as having a reasonable expectation of recouping any of the profits foregone by its low pricing conduct. Accordingly, the Bulletin notes that instances of true predatory pricing are usually rare and would be limited to markets with specific structural characteristics that allow the alleged predator to increase prices without fear of encouraging effective competitive entry in response.

The Bulletin proposes a *two-stage screening exercise* to determine whether these requisite structural characteristics are present. Because of the emphasis on market structure, elimination of a competitor or evidence of intent to do so generally would not be sufficient on their own to cause the DIR to exercise enforcement.

The *first stage* of the screening process is an *assessment of the degree of short-run market power possessed by the alleged predator*. This includes an examination of whether the entry and exit conditions in the relevant market might permit that firm to recoup losses caused by predatory pricing. The Bulletin makes it clear that, if such market conditions do not exist (that is, the alleged predatory firm is presumably only harming itself and not the alleged victim of predation in the market), the matter will not be pursued.

The second stage of the screening process examines *the pricing policy of the alleged predator in relation to cost information* and would come into play only if the Stage I market conditions have been satisfied.

The first task in Stage I is to define, through objective measures, the market power of the alleged predator. As a preliminary step, the relevant product market would be defined by examining both current and potential substitutes for the product whose prices are being examined. The geographic market would be established by an examination of consumer options for relocating their purchases in the event of significant price increase. For a predatory pricing inquiry to proceed, the alleged predator must have sufficient short-run market power to restrict output and raise prices through unilateral conduct.

The principal measures proposed by the Bureau for examining market power are the market share of the alleged predator; the measures of concentration in the relevant industry; the overall number and distribution of firms serving the relevant market; and the volatility of market shares of these firms. As a rule of thumb, the Bureau has proposed that the alleged predator must have at least a 35 percent share of the relevant market and be at least twice as large as its next largest competitor.

The Bureau has indicated that it would be unlikely to pursue any action against unilateral pricing conduct of a firm falling below these two thresholds. On the other hand, if a firm meets these criteria, the Bureau would still consider

other factors before determining whether to pursue the matter. As indicated by the Bulletin, such factors include the history and practices of the alleged predator, its overall size and financial strength, and any special advantages resulting from government intervention in the marketplace. Consideration would be given to whether the alleged predator is an incumbent firm or a new entrant.

In the Bureau's view, incumbent firms are more likely to engage in predatory pricing as they are more likely to have the incentive and ability to predate than do smaller entrants and may be in a better position to identify strategies to disadvantage their rivals.

This emphasis on relative market power as the initial screening device is consistent with the traditional notion that true anti-competitive low pricing will be practiced only by a significant incumbent firm to keep out vigorous new firms into its established market. In this initial stage, much will turn, therefore, on the definition of the relevant market. If the geographic and product markets are relatively narrowly defined in practice, this initial screening measure may persuade the Bureau to ignore low pricing behaviour by dominant or multi-product firms entering new markets with the objective of dominating those markets.

The second aspect of the first stage is an examination of conditions of entry and exit in the relevant market. The Bulletin notes that, for low pricing not to be a concern, the market must display the prospect of "effective entry," that is, not simply theoretical entry but whether "timely, sufficient" entry is likely. When both effective entry and exit are easy, the Bulletin contends, the initial low price behaviour of a firm's short-run market power will not be viewed as a threat to the competitive process. The following factors would be examined by the Bureau to determine whether the conditions for effective entry are present in that particular market.

(1) Speed of entry: This is regarded as the essential element for effective entry and the major focus of the Bureau's analysis. Speed of entry is defined to be the time required between identifying a business opportunity and selling in the relevant market. The Bulletin proposes that, as a general rule, a market is not exposed to effective entry if the minimum time required to enter exceeds 18 months. Accordingly, entry that would take several years

to accomplish would not deter or prevent “supra-competitive pricing” by a firm possessing market power once its low pricing policy achieved its short-term results of reducing competition.

(2) *Sunk costs*: These are investments the value of which could not be recovered in the event of business failure because they are either highly specialized or are not liquid. High sunk costs increase the financial risk of entry and reduce the expected short-run profitability upon entry.

(3) *Economies of scale and scope*: Economies of scale refer to the reduction of unit costs from increased volume of a firm’s output. Economies of scope refer to reduction of unit costs through the production of products jointly rather than individually. The two concepts are, in practice, inseparable for large multi-product firms. The presence of these elements is considered essentially a supplementary factor shading the analysis in the favour of conclusion that barriers to entry exist if there is evidence of low speed of entry and/or high sunk costs.

However, it is hard to see why the Bureau would consider the presence of economies of scale and scope as evidence of predatory pricing behaviour. Predatory pricing laws have been criticized for discouraging pricing innovations which would make a firm more efficient through economies of scale and scope. Predatory pricing laws are not supposed to penalize efficient firms for undercutting the prices of inefficient firms, particularly where excess capacity exists or where efficiency increases with a firm’s level of output.

In a technical sense, economies of scale and scope create a barrier to new entry. To become an efficient competitor, the new entrant must be capable of achieving the production volumes of the largest and hence the most efficient firm in the market. But, *if* in achieving its market foothold, the new entrant incurs higher unit costs than the established competitor, and the established competitor’s output decreases while its unit costs increase, and *if* the Bureau’s approach to predatory pricing inhibits incumbent pricing that may force out the less efficient new entrants, it is possible that industry efficiency and consumer welfare will end up in second place to the objective of increasing the number of visible competitors in the market.

Two examples are provided to explain why economies of scale and scope might count against an alleged predator. First, the Bulletin suggests that large-scale projects may require time-consuming plant construction that goes well beyond the period required by the predator to recoup any losses incurred from its predatory behaviour. However, if the new plant is more efficient than the predator's plant at a given level of output using average prices over the predation and post-predation time periods, the manner in which construction or start-up costs are accounted for should not matter. The more efficient plant should be built if it can produce at costs below that average price. If up-front accounting costs do matter, this arguably reflects more on imperfections in financial markets and accounting techniques than on possible market failure through predatory pricing.

The Bulletin also suggests that the entry may not be effective because of difficulties in overcoming brand loyalty. Arguably, brand loyalties have nothing to do with economies of scale and scope; rather, they represent imperfections of consumer information or simple consumer unpredictability or irrationality on the demand side of the market as opposed to the supply side (which side is more germane to determination of predatory pricing).

It is possible that the Bureau may reconsider the weight attached to economies of scale and scope in its Stage I analysis.

Finally, several other factors are mentioned as possible impediments to effective entry including institutional (patent, tariff or regulatory) barriers, established contractual arrangements of incumbent firms, and control over inputs by incumbent firms. The Bulletin also mentions that a firm with market power might signal to potential competitors that the market is unprofitable by pricing conduct, thus discouraging interest in the market.

In the event that this Stage I analysis reveals "a potential danger of effective predation," the Bureau would proceed to the second stage which involves an examination of price-cost relationships. The Bulletin emphasizes that no single price-cost test or criterion would be employed. In a restatement of the jurisprudence with respect to unreasonably low prices, the Bulletin suggests that whether certain prices are predatory depends on factors such as the duration of the period in which the low prices are maintained, whether they are adopted unilaterally or as a response to pricing policies of competing firms, and the underlying intent of the alleged predator.

Three general rules (again derived from the jurisprudence) are presented. First, a price at or above the average total cost incurred by the alleged predator is unlikely to be regarded as predatory. Secondly, a price below the average variable costs of the alleged predator is likely to be treated as predatory, unless there is clear justification. And thirdly, a price below the alleged predator's average total cost but not lower than its average variable cost (the "grey range") may or may not be treated as predatory depending on the circumstances.

These circumstances could include the intent of the pricing policy, the costs and financial weakness or strength of the target firm(s), the feasibility of re-entry of the market indicated by the Stage I analysis, the existence of excess capacity, and general demand conditions prevailing in the market.

In adopting this modified form of a variable cost threshold for anti-competitive prices, the Bureau has unfortunately provided only thin guidance to industry on how to determine variable costs. How the Bureau would determine these costs has also been left unclear since it would appear that the only way the Bureau could obtain useful evidence of the alleged predator's variable cost would be through the exercise of formal investigatory powers following the initiation of a formal predatory pricing inquiry under the *Competition Act* rather than in the course of a pre-inquiry screening analysis.

The Bulletin notes that variable costs include costs that may be varied with levels of output, including labour, material energy, promotional allowances and use-related plant depreciation.

With multi-product firms, of course, the exercise of identifying direct and indirect variable costs with particular product lines and output changes has proven to be a very difficult and often arbitrary exercise. No specific guidance is provided on the appropriate principles of common or joint cost allocation. The Bulletin offers no guidance with respect to the period of time over which the variability of particular input costs with levels are to be determined. Over a sufficiently longer period of time, of course, all costs are variable. Needless to say, no conventional cost-accounting framework provides a guarantee that all costs can readily be causally related to variation in a particular product line's output, even though all budgeted costs vary to a degree with the budgeted revenues of a firm.

The Bulletin does indirectly suggest that fixed costs include costs associated with investment in plant and machinery and fixed assets. On the other hand, the Bulletin suggests that “use-related plant depreciation” is a variable cost. It is hard to determine the Bureau’s boundary line between fixed and variable costs. In practice this distinction can vary considerably according to the parameters of the cost analysis employed.

Moreover, many firms’ physical facilities and machinery can be incremented easily within the 18-month period established by the Bulletin for assessing likely effectiveness of entry. Such plant investment may be traced to specific product lines or addition of volumes produced in existing product lines. Are these costs fixed or variable?

To cloud the picture further, the Bulletin suggests that its cost analysis would also be based on “reasonably anticipated rather than actual variable costs.” The question arises then as to whether the Bureau, for whatever time frame it selects to analyze the predatory behaviour, might unilaterally impose adjustments to the existing or recorded cost structure of the alleged predator. These adjustments might be based on, for example, inflation and increased excess capacity caused by a loss of market share from successful entry by the alleged victim of the predator’s pricing conduct or, alternatively, the Bureau might cost assets at their current replacement cost as opposed to their recorded historical and depreciated cost.

Finally, the Bureau suggests that its price-cost analysis need not be restricted to a static analysis. It may take into account the possible future cost structure of incumbent firms if additional plant capacity is built in response to entry by the alleged victim of predatory pricing. Consequently, the Bulletin suggests that the timing of plant increments in relation to new entry would be a relevant consideration. The application of hypothetical costs from yet to be built or newly on-stream capacity increments, particularly where production technology is changing, may further complicate the price-cost analysis. This, in turn, will further reduce the ability of business planners to anticipate the reaction of the Bureau to low pricing conduct should a matter reach Stage II of the Bureau’s preliminary analysis.

The Bureau has also indicated that, in any event, it does not intend to be bound to the results of whatever cost test it applies. Where prices are below the average total cost but above the average variable cost (the most likely

outcome), the Bureau would take into account the surrounding circumstances. These could include the intent of the pricing policy, the costs and financial weaknesses or strengths of the target firm(s), the feasibility of re-entry to the market as revealed in the Stage I analysis and the existence of excess capacity and general demand conditions prevailing in the market.

To ensure that the potential for certainty is clouded, the Bulletin then concludes that the inferences drawn will depend directly upon the apparent purpose of the low pricing and its reasonableness in light of the facts. (p. 12)

Thus, having apparently rejected the subjective, or intention-driven, aspect of the offence in favour of an objective, structural- and cost-driven analysis at the outset of the Bulletin, the Bulletin reintroduces intention as a critical swing variable in determining whether to pursue cases in the grey range of the Stage II analysis.

C. CONSPIRACY

1. Introduction

The core of the *Competition Act* is the criminal prohibition against agreements or arrangements to lessen competition unduly. The Act states:

- 45.(1) Every one who conspires, combines, agrees or arranges with another person
- (a) to limit unduly the facilities for transporting, producing, manufacturing, supplying, storing or dealing in any product,
 - (b) to prevent, limit or lessen, unduly, the manufacture or production of a product or to enhance unreasonably the price thereof,
 - (c) to prevent or lessen, unduly, competition in the production, manufacture, purchase, barter, sale, storage, rental, transportation or supply of a product, or in the price of insurance on persons or property, or
 - (d) to otherwise restrain or injure competition unduly,
- is guilty of an indictable offence and liable to imprisonment for a term not exceeding five years or to a fine not exceeding ten million dollars or to both.

Section 45(2) is intended to clarify the concept of undue effect by not requiring that complete elimination of competition in the relevant market was the result of the agreement or the object of the parties:

(2) For greater certainty, in establishing that a conspiracy, combination, agreement or arrangement is in contravention of subsection (1), it shall not be necessary to prove that the conspiracy, combination, agreement or arrangement, if carried into effect, would or would be likely to eliminate, completely or virtually, competition in the market to which it relates or that it was the object of any or all of the parties thereto to eliminate, completely or virtually, competition in that market.

Introduced with the 1986 revision, sections 45(2.1) and 45(2.2) are intended to prevent judicial interpretation of the offence as requiring proof of communication among the parties to prove the existence of an agreement, and requiring proof that the parties specifically intended that the agreement would lessen competition unduly:

(2.1) In a prosecution under subsection (1), the court may infer the existence of a conspiracy, combination, agreement or arrangement from circumstantial evidence, with or without direct evidence of communication between or among the alleged parties thereto, but, for greater certainty, the conspiracy, combination, agreement or arrangement must be proved beyond a reasonable doubt.

(2.2) For greater certainty, in establishing that a conspiracy, combination, agreement or arrangement is in contravention of subsection (1), it is necessary to prove that the parties thereto intended to and did enter into the conspiracy, combination, agreement or arrangement, but it is not necessary to prove that the parties intended that the conspiracy, combination, agreement or arrangement have an effect set out in subsection (1).

The offence does not apply to arrangements that relate only to the following, unless there is proof that the agreements actually lessen competition in respect to prices, quantity or quality of production, markets or customers, or channel or methods of distribution:

(a) the exchange of statistics;

- (b) the defining of product standards;
- (c) the exchange of credit information;
- (d) the definition of terminology used in a trade, industry or profession;
- (e) cooperation in research and development;
- (f) the restriction of advertising or promotion, other than a discriminatory restriction directed against a member of the mass media;
- (g) the sizes or shapes of the containers in which an article is packaged;
- (h) the adoption of the metric system of weights and measures; or
- (i) measures to protect the environment.

Export agreements are also exempted unless they result in a reduction in the real value of exports of a product; restricted entry or expansion in an export business; or undue restriction of competition in the supply of services facilitating Canadian exports.

In my view, the conspiracy prohibition applies now to regulated passenger transportation services to the same extent that it would apply if the applicable regulatory structures were removed. Note, however, that the application of the conspiracy provisions is limited in relation to the Crown and its agenda to the same extent as are all other provisions of the Act.

The conspiracy jurisprudence is extensive. Over the last decade, convictions or pre-conviction prohibition Orders (a substitute to a plea bargain specifically provided for by the Act) have generally been obtained when charges have been laid. The circumstances giving rise to charges generally have involved an actual or constructive agreement among all major suppliers in a market not to compete with respect to an important aspect of competition (price, product quality, product choice); or some disciplinary mechanism applied by wayward incumbents to innovative new entrants (for example, withdrawal or refusal of a group service needed to function as an effective competitor).

2. Prohibition Agreements

Important prohibition agreements with respect to the used household goods moving industry and the Western Canada for-hire trucking industry have effectively eliminated the use of tariff bureaux as a cartelizing mechanism

and have introduced price competition for the first time in these sectors. These prohibition agreements have been influential in re-orienting the Bureau's conspiracy law enforcement policy. Instead of seeking convictions and large fines, the Bureau is leaning towards obtaining assurances that affirmative steps would be taken more quickly by the accused to open up the market and, to a degree, offset some of the negative affects of the agreements.

The used household goods prohibition Order, entered into in December 1983, requires the winding up of the Canadian Household Goods Carriers' Tariff Bureau and prohibits several restrictive practices by van lines including:

- rate fixing;
- adoption of fighting brands (that is, lower priced products targeted on a competitor's customers);
- using cost studies to coordinate rates;
- standardizing or limiting products;
- coordinating van utilization or requiring that products be supplied to discipline competition;
- generally using a van line's market power as franchisor to discipline the competitive conduct of van line members or to coordinate competition among van line members.

Periodic compliance reports were required to be filed with the Bureau and the court.

The Western trucking prohibition Order was entered into in April 1988. The inquiry on which charges were based commenced in the 1960s. The accused trucking companies (18 in all) and the Western Transportation Association (WTA) were prohibited from agreeing with any other motor carrier to:

- (a) fix or coordinate single line rates in the market;
- (b) enforce adherence to single line rates in the market as published by the WTA or any other motor carrier tariff bureau or motor carrier industry association;

- (c) develop, adopt or use any policy, plan or program to respond to any motor carrier operating in the market;
- (d) attempt, directly or indirectly, by threat, promise or any like means, to influence upward or to discourage the reduction of, the price at which any other motor carrier supplies or offers to supply single line services in the market; or
- (e) restrict or impede in any manner the entry of any motor carrier competitor or potential motor carrier competitor in the market.

The trucking companies were also prohibited from using the Western Transportation Association or any other industry association, to signal adoption of single line rates unless notice was given to the public by other means.

The WTA and its officers and employees were prohibited from:

- (a) initiating tariff rate proposals, docketing their own tariff rate proposals or making any recommendations whether to adopt, reject or otherwise dispose of tariff rate proposals before the WTA applicable to the market;
- (b) initiating or developing any collective response among the members of the WTA to rates proposed or changed by any motor carrier operating in the market.

However, the Order does not extend to agreements on interline rates.

Taken together, the prohibition Orders set out with some certainty the types of tariff-bureaux conduct and rate arrangements in the transportation sector that the Bureau will likely challenge.

3. Constitutionality

The effectiveness of the conspiracy prohibition has recently been seriously called into question by a September 1990 decision of the Nova Scotia Supreme Court which found the prohibition, as drafted, in violation of the *Canadian Charter of Rights and Freedoms* and of no force as a consequence.

The case arose from charges brought in Nova Scotia against 12 pharmaceutical firms. The charges alleged that the firms had illegally conspired between January 1974 and June 1986 to lessen competition in the sale

and provision of prescription drugs and dispensing services. The accused brought a motion to have section 45 ruled invalid as being contrary to sections 7, 11(a) and 11(d) of the *Canadian Charter of Rights and Freedoms*.

The first argument of the accused was that the *mens rea* (subjective intent) required to obtain a conviction violated section 7 (the right to life, liberty and security of the person deniable only in accordance with the principles of fundamental justice) and section 11(d) (presumption of innocence of an accused until proven guilty at a fair public hearing).

The Court found that the *actus reus* of the offence contained two elements:

- (a) an agreement to which an accused was party; and
- (b) the agreement, if implemented, would have the effect of limiting competition unduly.

The Court concluded that, in restricting proof of *mens rea* only to the first element, the provision violated sections 7 and 11(d) of the Charter.

The accused also contended that the undueness element of the offence was so vague as to deny their section 11(a) and (d) rights to a full answer and defence, and a fair trial. The Court agreed, stating:

As indicated in Reference Re s.193, *supra*, the test, for determining whether or not a law is vague, as stated by Chief Justice Dickson and quoted earlier, is whether a person is capable of knowing, in advance, with a high degree of certainty, what conduct is prohibited and what is not. In my opinion, the virtual monopoly definition, of Mr. Justice Cartwright, provided some degree of certainty, but Parliament has eliminated that definition. The Crown says it is a question of degree, and that evidence of lessening competition was so extensive that it would be shown to be undue, but this does not answer the "knowing in advance" portion of the Reference Re s.193 test.

Having so found, it was easy for the Court to conclude that it would not be possible for the Crown to provide sufficient information in an indictment to ensure a fair and full trial.

The Court also addressed whether section 1 of the Charter could sustain these provisions, notwithstanding the violations of guaranteed rights. Section 1 states:

The *Canadian Charter of Rights and Freedoms* guarantees the rights and freedoms set out in it subject only to such reasonable limits prescribed by law as can be demonstrably justified in a free and democratic society.

In reviewing the cases considering section 1, the Nova Scotia Court quoted, in particular, from a decision of Mr. Justice Graham of the Ontario Court of Appeal in *R. v. Seaboyer* (1987), 20 O.A.C. 345 (O.A.C.) where he stated:

I think it would be a most unusual result that a law which offends s. 7 in that it deprives an accused person of making full answer and defence could ever be found "to be demonstrably justified in a free and democratic society."

In conclusion, the Court stated: "This is not one of the rare or exceptional cases such as a war or an epidemic where a s. 7 violation can be justified under section 1. Nor is the vagueness and uncertainty, contained in this section, a limitation prescribed by law."

Having come to these conclusions, the Nova Scotia Court then struck down the present sections 45(1)(c) and 46 as being contrary to the Charter. The Court also quashed the indictment in its entirety even though it dealt with events which took place before the passage of the Charter in 1982.

The Crown is appealing the decision. The Bureau has taken the position, as it normally does in such constitutional disputes, that the conspiracy law will continue to be applied in all other provinces until this case is determined. However, the Bureau's position may have to change if the Nova Scotia Court of Appeal upholds the decision or if a federal court in a different case reaches the same conclusion.

This decision again highlights the difficult constitutional underpinnings of federal competition law under the current constitutional structure and will certainly cause a reconsideration of structuring key elements of the law as criminal offences. It is worth noting that federal government proposals

for competition law reform dating from the mid-1970s favoured the shift of all criminal matters, including conspiracy, to a civil law context to be adjudicated by an expert tribunal.

The case will undoubtedly be resolved by the Supreme Court of Canada. This will probably take at least two years.

This decision was a shock to what was thought to be the firmest foundation of Canadian competition law and will likely test the commitment of the federal government to a federal competition law.

D. MERGERS

1. Brief History

The 1986 revisions to the *Competition Act* included a dramatic change to the general merger law, replacing a clearly ineffective criminal prohibition with a complex standard for merger review by the Competition Tribunal.

However, since the new merger regime took effect four years ago, only one contested merger (that of Reservec/Gemini) has reached the trial stage and two others have been the subject of yet-to-be adjudicated applications (Quebec meat renderers and British Columbia south mainland newspapers). A much larger number of mergers challenged by the Bureau have resulted in agreement between the merging parties and the DIR to restructure the merger to remove the aspect of the transaction that would have caused a substantial lessening of competition in the relevant markets. In consideration of this, the DIR did not make application to the Tribunal.

Typically, these agreements have involved the divestiture of certain assets or product lines to third parties to increase overall competition in the industry and to reduce market power in the most concentrated submarkets supplied by the merging parties.

To date, only horizontal mergers have been challenged. A general rule has developed over the first three years of enforcement. A merger would be challenged if it involved two of the three largest suppliers in a market and resulted in the merged company having a market share above 50 percent with no comparable competitive market share.

More frequently, the DIR has submitted a consent arrangement for Tribunal approval. In two cases, *DIR v. Palm Dairies Ltd.* (1986), 12 C.P.R. (3d) 540 (Comp. Trib) and *DIR v. Imperial Oil Ltd.* (unreported Comp. Trib. decision, February 6, 1990, CT-89-3), the Tribunal exercised its independence and rejected the original proposals on the basis that they did not sufficiently correct the lessening of competition caused by the merger.

Since the new merger law had generated very limited true jurisprudence, particularly with respect to what constitutes a substantial lessening of competition, in November 1990, the Bureau published proposed merger guidelines to increase the predictability of the law. These guidelines were finalized in March 1991. They are discussed briefly below. Rather than providing a general survey of all merger resolutions, the following section reviews the Director's approach to the PWA/Wardair merger and the resolution of the Reservec/Gemini CRS merger case.

2. Merger Law

The Act defines a merger broadly as the acquisition or establishment, direct or indirect, by one or more persons, whether by purchase or lease of shares or assets, by amalgamation or by combination or otherwise, of control over or of significant interest in the whole or part of a business of a competitor, supplier, customer or other person (section 91).

The Bureau considers a "significant interest" to be less than a controlling position but entailing the actual or potential ability to influence materially the firm's economic behaviour.

A reviewable merger is an actual or proposed merger which prevents or lessens, or is likely to prevent or lessen competition substantially in a market (section 92(1)).

The Tribunal may not make a finding of substantial lessening of competition solely on the basis of evidence of concentration or market share (section 92(2)).

The Act provides a non-exhaustive discretionary list of factors for assessing the impact on competition:

- (a) the extent to which foreign products or foreign competitors provide or are likely to provide effective competition to the businesses of the parties to the merger or proposed merger;
- (b) whether the business, or a part of the business, of a party to the merger or proposed merger has failed or is likely to fail;
- (c) the extent to which acceptable substitutes for products supplied by the parties to the merger or proposed merger are or are likely to be available;
- (d) any barriers to entry into a market, including:
 - (i) tariff and non-tariff barriers to international trade,
 - (ii) interprovincial barriers to trade, and
 - (iii) regulatory control over entry,and any effect of the merger or proposed merger on such barriers;
- (e) the extent to which effective competition remains or would remain in a market that is or would be affected by the merger or proposed merger;
- (f) any likelihood that the merger or proposed merger will or would result in the removal of a vigorous and effective competitor;
- (g) the nature and extent of change and innovation in a relevant market; and
- (h) any other factor that is relevant to competition in a market that is or would be affected by the merger or proposed merger.

The Tribunal may not issue a remedial Order if it finds that the merger is likely to bring about efficiency gains that will be greater than, and offset, the effect of any lessening of competition and these efficiency gains would not likely be attained if the remedial Order were made (section 96).

Finally, certain joint ventures of limited duration that meet conditions set out in the Act are also exempted (section 95).

The merger guidelines emphasize that the primary focus in determining an effect on competition is the ability to exercise a greater degree of market power, with the price dimension of competition being the dominant concern. Relevant price increases may either be unilateral or the result of "interdependent behaviour." That is, the possibility that the merger might result in

collusion or even conscious parallelism would be taken into account. This concern over an enhanced ability to collude has been supported in both the Reserve/Gemini and Imperial Oil/Texaco cases.

The guidelines provide that a substantial effect on prices would be one that is materially greater over a substantial part of the effective market and that new or increased competition would not likely be eradicated within two years.

3. Market Definition

Definition of the relevant market is critical in merger assessment. The draft guidelines provide, after an extensive discussion, the following conceptual definition: . . . a relevant market is defined as the smallest group of products in the smallest geographic area in relation to which sellers could impose and maintain a significant and non-transitory price increase above the level that would likely exist in the absence of the merger. In most cases the Bureau considers a 5 percent increase to be significant, and a one year period to be non-transitory.³

This hypothesis will certainly be hard to apply. Consequently, the merger guidelines present a lengthy list of more subjective market definition criteria:

- (a) the views and behaviour of buyers of the product;
- (b) the views of other competitors in the same business;
- (c) the functional substitutability of products in the market;
- (d) the physical and technical characteristics of the product;
- (e) the costs a buyer would incur in switching to a different product;
- (f) price relationships or relative price levels between two products;
- (g) cost of adapting or constructing facilities to sell the product; and
- (h) whether there exists a second-hand, reconditioned or leased product market.

With respect to geographic market definition, the evaluation criteria set out in the guidelines are:

- (a) the views of buyers in the marketplace;

- (b) the switching costs of moving from one producer's product to another;
- (c) transportation costs as an impediment to sourcing product outside a defined area;
- (d) any local setup costs for a new entrant;
- (e) inherent characteristics of the product which could be relevant, such as perishability;
- (f) review of price movements in different geographic areas and any evidence of correlation;
- (g) the shipment patterns of the product in the past; and
- (h) foreign competition.

4. Statutory Evaluation Criteria

The guidelines also address each of the statutory factors listed above. The objective is to suggest thresholds below which it is unlikely a merger would ever be challenged. For example, with respect to the critical concepts of market share and concentration, it is proposed that the merger would not be challenged where the post-merger market share was less than 35 percent, or where the post-merger market share of the four largest firms in the market was less than 65 percent and the market share of the merged firm was less than 10 percent.

Of the remaining factors, barriers to entry, effective competition remaining and failing firm are the most important.

If entry were to take place on a sufficient scale to ensure that a price increase could not be sustained for more than two years, the merger would not be challenged.

Where the pre-merger level of effective competition would remain, the merger would not be challenged.

The guidelines present a detailed discussion of what constitutes a failing-firm situation. To accept this as a ground for not challenging the merger, the Bureau would require identification of the time within which the firm

might become insolvent, a reasonable search for a purchaser prepared to pay more than the firm's breakup value, and provision of detailed supporting financial information. The guidelines state that the Bureau needs about 60 days to determine whether a firm is indeed failing and may require a third party to find a buyer.

The guidelines also raise two new factors: market transparency and transaction value and frequency. With respect to market transparency, the guidelines state:

... where a merger raises concerns that it may be likely to facilitate interdependent behaviour, the extent of transparency in the relevant market will ordinarily be assessed. Transparency in this context connotes information that is readily available in the market about competitors: prices; levels of service; innovation initiatives; product quality; product variety; level of advertising; etc.. In general, as the level of transparency in a market decreases, coordinated behaviour becomes increasingly difficult, because firms find it harder to detect and retaliate against secret discounts and other deviations from interdependent situations. (p. 42)

The guidelines also talk about schemes which increase transparency such as delivered pricing systems, product standardization, exchanges of information through trade associations, etc., public disclosure of information by buyers or government sources, and "meeting competition" contractual terms.

The second new factor — transaction value and frequency — is also related to mergers where independent behaviour may be a concern. It states:

Interdependent behaviour often becomes increasingly difficult as the frequency and regularity of sales of the relevant product decrease, and as the value of each sale increases. (p. 43)

Although the guidelines do not state how important these two criteria are, they both relate to the interdependent behaviour or tacit collusion on which the DIR might challenge a merger.

5. Procedure

The guidelines indicate that the DIR should be able to make a preliminary determination within three weeks as to whether there is a serious question. If there is, the guidelines state:

... a determination can be made of whether a merger prevents or lessens competition substantially within eight weeks after the parties have provided all requested information. This period of time is required in order to review this information, and to gather and review information provided by customer, suppliers, competitors, experts, others in the industry and government departments that have information pertaining to the market(s) in question. Where information is not provided upon request by merging parties or others, the Director may seek to exercise the formal powers provided under section 11, 15 or 16 of the Act. (p. 60)

The guidelines also indicate that where there is still a significant issue after this period, which presumably means that the DIR has concerns about the merger, the DIR can take up to a further four months to decide whether to refer the matter to the Tribunal.

6. PWA/Wardair Merger

The first merger clearance accompanied by a fairly detailed set of reasons related to the 1989 merger of PWA and Wardair.

After an investigation, the DIR announced on April 14, 1989 that he believed Wardair to be a failing firm and that, in those circumstances, the merger would be challenged only if a viable third party came forward and was willing to acquire Wardair. In the week that followed the Director's announcement, there were other expressions of interest in Wardair, including a possible joint offer between American Airlines, the large U.S. based carrier, and certain Canadian partners. Finally, however, no offer was forthcoming by any third party.

It is apparent from the background information provided by the DIR, that the acquisition raised serious concerns. The DIR viewed Wardair as a vigorous and effective competitor. The background information document states:

Overall, in markets where it participated, Wardair was often the price leader, offering effective competition to the two established national carriers in all classes of service.⁴

The DIR also was concerned that the acquisition would threaten effective competition in the domestic airline service industry. The background document stated that most scheduled airline service in Canada would be provided by Air Canada and Canadian Airlines International Ltd. and that consumers would have little countervailing power to deal with such a “duopoly.”

The DIR indicated there were other reasons to expect that the merger would lessen competition. In particular, he noted that the stock market reaction, which had led to higher stock prices for both Air Canada and PWA, was a sign that investors expected increased profits for the rivals of Wardair. Also, he cited statements by officials of Wardair and PWA that the merger would reduce capacity in the industry and restrict the availability of discounted seats.

The DIR was also concerned with barriers to entry to the domestic airline industry in the short to medium term. He indicated that:

Recent economic analysis of the industry indicates that the industry reflects the presence of economies of scope which give scheduled carriers offering a range of fare types cost advantages over scheduled carriers offering one-fare, one-class service, and over chartered carriers that can offer only restricted fares with travel conditions on return dates and itinerary changes.⁵

The DIR referred to the preference of business travellers for an airline with an extensive network and frequent flights, the loyalty generated by frequent flyer programs, and the preference for good commuter connections made easier by the two largest carriers.

The DIR also cited the limited ability of Transport Canada to provide landing rights to new carriers at Pearson International Airport in Toronto. The DIR stated that:

At Pearson International Airport, the most critical public resource at the moment is the take-off and landing times at peak hours. These are

currently limited to 70 per hour. Accommodating new entrants at this juncture will be difficult. Nevertheless, Transport Canada has recently reaffirmed to the Bureau that every effort will be made to do so.⁶

As a final barrier to entry, the DIR mentioned the restrictions on foreign ownership which allow non-Canadians to hold a maximum 25 percent voting interest in a Canadian scheduled carrier. He also noted that cabotage rules prohibit foreign carriers from offering point-to-point scheduled service within Canada. The Canadian cabotage rule is similar to that in the United States.

Nevertheless, the question of the failing-firm factor under the merger law remained. The DIR noted that there were two significant issues to be assessed in considering the failing-firm issue: first was the extent to which the failure is, in fact, likely to occur and, secondly, whether there are alternatives to the merger that would be less restrictive of competition.

With respect to the first factor, the DIR stated:

A firm that is facing certain and imminent financial failure will cease to exercise any competitive influence in the market after its failure. Therefore, the loss of this influence in the marketplace cannot be attributed to the merger.⁷

The accounting firm of Peat Marwick, retained by the DIR to review the financial evidence provided by Wardair, reviewed the options which might have prevented the failure of Wardair. These included deferring principal payments, loans on existing fixed assets, the sale of a minority interest and reversion to a chartered carrier status. The Director's background statement indicates that none of these are attainable or workable in the circumstances of Wardair.

With respect to alternatives, the DIR considered third-party buyers and also liquidation of the failing firm. The DIR concluded that the most likely result would be the withdrawal of Wardair's assets from the Canadian market. The background statement indicates:

There now exists considerable excess capacity in the market and Wardair's A310 aircraft are not compatible with the fleets of other airlines in Canada.⁸

Having weighed all of these factors, the DIR concluded that the significance of the failing-firm factor outweighed the negative assessment of the other competition-based factors in the Act.

It is evident that the Director's decision was a difficult one. The effect on competition of the withdrawal of Wardair was immediate and palpable. Fares went up; fare choices declined.

The failure of Wardair raises serious questions more about the viability and effectiveness of airline deregulation in Canada in the context of remaining regulatory entry barriers effectively controlled by Transport Canada and less with respect to the effectiveness of the merger law.

These government-controlled entry barriers are principally:

- (1) *the statutory foreign investment ceiling* of 25 percent which reduces the access of new and existing carriers to the capital market and severely reduces the effective market for control of larger existing carriers. In particular, this investment limit creates a significant disincentive to the transfer of technology and managerial skills from non-Canadian air carriers to Canadian air carriers;
- (2) *landing and takeoff slot allocation practices* of airport management which "grandfather" incumbent allocations, effectively control new entry to airports through incumbent-dominated advisory committees, and do not ration supply on the basis of the marginal value of slots; and
- (3) *limitation of foreign, and particularly U.S., air carrier access* to the Canadian market through treaty arrangements.

The Director's background statement identifies three reasons why deregulation may not have been as successful as expected when it was introduced in the early part of this decade. The first is the possibility that economies of scope are a significant barrier to entry into the scheduled airline business. When the United States and Canada deregulated airlines, the common view was that economies of scale were small in the industry and that economies of scope were not significant. Since then, it has become clear that economies of scope are much more important than previously thought. Secondly, restrictions on foreign ownership and foreign operations in Canada give the federal government the means to

inject competition in the domestic market if it wishes to do so. Whether Transport Canada will relax these requirements, however, remains to be seen. Finally, the analysis points to a third factor, access to hub airports. Since this is controlled by committees made up of incumbent carriers and Transport Canada officials, it is also seen as a serious barrier to entry.

The National Transportation Agency found that the Wardair acquisition by PWA was not against the public interest. Its decision was based largely on the financial health of Wardair, as was the Director's decision. Indicating that it was in no way interfering with the DIR's review, the NTA stated that the purpose of its review was different from that being conducted by the DIR under the *Competition Act*. This is apparently the first time the NTA has explicitly recognized the separate jurisdiction of the DIR to review the same merger transaction.

7. Reservec/Gemini Merger

On March 3, 1988, the DIR applied to the Competition Tribunal for an Order ordering Air Canada (AC) and Canadian Airlines International Limited (CAIL) to dissolve their limited partnership instituted to combine the operations of the Reservec and Pegasus Computer Reservations Systems into a single system known as Gemini.

Computer Reservations Systems (CRSs) are an increasingly important element in the distribution and sale of airline passenger seats to travel agents and to the travelling public. The systems distribute information on schedules, fares, rules and seat availability to subscribers (usually travel agents) for the airlines which are hosted on, or participate in, the system. This information is distributed electronically through a Computer Reservation Terminal (CRT) which is sold or leased to the subscriber and is located on the subscriber's premises.

The Director's application stated that before the merger, Air Canada's Reservec distributed information to approximately 2,900 travel agencies on behalf of 50 airlines, railways and car rental agencies, 3,000 hotels and 16 tour wholesalers. The application indicated that Reservec was the dominant CRS in Canada, holding about 72 percent of the CRS market as measured by travel agent locations.

Pegasus was developed by Canadian Pacific Airlines before its 1987 merger with PWA. Pegasus entered the Canadian market in 1984 and, according to the Director's application, "introduced some innovative features, providing competition for Reservec." The application claimed that between 1984 and 1987, Pegasus established its system in approximately 720 travel agencies providing information on and to 60 airlines, 14 car rental agencies, 3,000 hotels and tour wholesalers. The application stated that Pegasus was the second largest CRS in Canada, holding approximately 18 percent of the CRS market as measured by travel agent locations.

On June 1, 1987, AC and CAIL (the airline resulting from the CP/PWA merger) took steps to merge the Reservec and Pegasus systems through a limited partnership in which each of the parties received partnership units and other consideration reflecting the proportion of assets contributed to the partnership.

The DIR submitted that there are no effective substitutes to a CRS and that prior to the merger there were three other competitors to Reservec and Pegasus, including Sabre, a subsidiary of AMR Corporation which also owns American Airlines. Sabre had entered the Canadian market in 1983 and, by June 1987, had approximately 10% of the market as measured by travel agent locations. The other two competitors noted are Apollo, a CRS operated by Covia Corporation which is owned by United Airlines, and Soda/System One, owned by Texas Air Corporation. The application claimed that the latter two CRSs have an extremely small presence in the market with a combined market of less than 1%. To establish the importance of CRSs in the supply of airline tickets, the application noted that travel agencies are now the primary means for airlines to distribute their product to the travelling public. Approximately 70% of the tickets sold by Canadian airlines are sold through travel agents and approximately 90% of all Canadian travel agencies use CRSs to make airline reservations and to print tickets. The other 30% of the tickets are sold by the airlines directly to the travelling public. In almost all cases, the application claimed, the airlines use CRSs to assist with sales.

Critical to the Director's theory of the case was the distinction between the CRS services available to a "hosted carrier," and the lesser grade of CRS services available to a "participating carrier." If a carrier is hosted, the CRS stores the carriers complete inventory information. In this case, the CRS provides the carrier with both an internal reservation and management

system to manage its inventory and an external reservation system to distribute its product to travel agents and consumers. Air Canada and Canadian Airlines International are now hosted with the Gemini system.

If an airline is a participating carrier, the CRS does not supply an internal reservation and management system but instead only lists the information on fares, schedules and seat availability which the participating carrier supplies. A participating carrier may choose not to supply all of its inventory so that certain classes of seats may not be displayed on the CRS in which the airline is participating.

The application alleged that a hosted carrier has a significant competitive advantage over participating carriers because of the completeness, accuracy and timeliness of information on seat availability from the CRS with respect to hosted carriers.

The application also stated that, for practical purposes, an airline can store its entire inventory in only one place, which means that it can participate in a number of CRS systems but can be hosted by only one. Many of the advantages of hosted carrier status, however, can be obtained by means of a direct access data link between the CRS and the data base of the participating airline. The application noted that there are several CRS vendors in the United States, all of which have a direct access link with carriers who are hosted in another CRS:

These links mean that these CRS vendors compete on the basis of what their systems can do and the price at which they do it rather than on the basis of exclusive control of airline inventory. In Canada, prior to October 31, 1987 there were no direct access links between the three largest CRS vendors in Canada namely Reservec, Pegasus and Sabre. On or about October 31, 1987 an electronic direct access link was established between Reservec and Pegasus, giving users of either Reservec or Pegasus last seat availability on Air Canada and Canadian Airlines International.

Last Seat Availability refers to the capacity of the CRS to call up for reservation seats held back by the airline from CRS booking of the airline. It is regarded by travel agents as an important competitive feature.

The principal grounds presented by the DIR in his application were:

- *Increased concentration:* Gemini's post-merger market share is calculated at 90 percent (of travel agent locations versus 10 percent for Sabre) and "has reduced the number of significant CRS competitors in Canada from three to two and in many non-urban areas has eliminated competition completely."
- *Increased barriers to entry:* The application contended that the superior service delivered by Gemini for AC and CAIL and their affiliated and aligned carriers, coupled with the current dominant position of these two carriers in the Canadian market, provides Gemini and its owners with the ability to block or frustrate the entry of competing CRSs by reducing the access of competing systems to timely and reliable information on the operations of AC and CAIL. The application also claimed the possibility of new entry into the CRS by a non-airline vendor is remote because of the substantial software and hardware development costs and the fact that airline vendors enjoy significant economies of scope because they must have a reservation system in any event.
- *Lack of availability of substitutes:* The application claimed that other sources of information such as manual reference to the *Official Airline Guide* and the use of the telephone to make airline reservations are too time consuming to be a practical alternative for most travel agents.
- *Effective competition remaining and removal of a competitor:* The application stated that the merger would reduce the effectiveness of Sabre as a competitor because Sabre, in the absence of a direct access link with Air Canada, CAIL and Gemini, would not be able to provide its travel agent subscribers with Last Seat Availability and other enhancements on AC and CAIL flights available through Gemini. The application also alleged that Sabre could be quickly neutralized by AC and CAIL if they exercised the market power they hold by reason of their dominant position in the airline market. For example, the application suggested that withdrawal of their participation in Sabre "would likely force Sabre to withdraw from the Canadian market because Sabre would then be providing a service without any booking fee revenues."
- *Impact on airline industry competition:* The application also alleged that the merger would likely entrench the dominant position of AC and CAIL in the airline industry at the expense of Wardair and potential new entrants in both the jet carrier and turbo-prop markets in Canada; it would also be

a detriment to U.S. and international carriers who compete in transport or international markets with AC and CAIL. In this regard, the application suggested that competing carriers which host or participate with Gemini may be subject to bias and other disadvantages which could severely inhibit their ability to compete. The suggested disadvantages include denial of access to the CRS, inaccurate loading of information, biased flight display ordering, and discriminatory booking fees. However, it should be noted that the application did not allege that any such practices have in fact taken place in Canada. The application essentially suggests that such activities would be more likely were the merger to take place.

The application therefore presented a number of important issues:

- To what extent should concentration in one market in transport be used in assessing the impacts of a merger in the production of a complementary product (CRS), and conversely what weight should the Tribunal attach to the possible lessening of competition in air transport that may result from the CRS merger?
- Is dissolution of the merger the only appropriate remedy? Could the possible anti-competitive impacts of the merger be sufficiently reduced if Air Canada's and CAIL's competitors were provided with direct access data links with those carriers' reservation systems as is done in the U.S.? To what extent is the viability of this option affected by the more concentrated nature of the Canadian air transport industry compared to the U.S. industry?
- What weight, if any, should be given to any increased potential for the abuse of dominant position in the market of the merger or related markets in determining whether the merger substantially lessens competition?
- Does the merger entail efficiency gains which are not likely to arise if the merger were dissolved and which, therefore, may exempt it from a remedial Order notwithstanding that it may lessen competition?

(a) Air Canada and CAIL Responses

The written responses of Air Canada and CAIL followed similar lines and disputed the Director's application in the following principal areas:

- market definition;

- efficiency gains and competitiveness in the CRS business; and
- impact on competition in the air transport business.

With respect to market definition, the respondents contended that in certain respects the appropriate market for CRS and airline services should include the domestic, transborder and international markets. In such a market, Gemini remains a minor player relative to Sabre, the dominant CRS in North America, and the principal remaining CRS competitor in Canada. The respondents also contended that measurement of market share according to travel agencies served (as was done in the Director's application) is not appropriate and leads to an overstatement of Gemini's share. Booked flight segments was proposed as the appropriate proxy for CRS revenues.

Several efficiency-related arguments were presented:

- The merger resulted in cost savings of \$15 million a year flowing from economies of scale (reduced facilities and operational duplication).
- The world market could sustain only a few CRS firms due to the requirement for large-scale technology and the inherent cost advantages of an airline-sponsored CRS. A minimum efficient firm size or "critical mass" for the North American and Canadian markets is at least that of Gemini, if not larger, and only CRSs sponsored by major airlines are likely to survive over the long term. Air Canada and CAIL, therefore, contended that the merger increased the competitiveness of Gemini against rivals such as Sabre which had already reached an efficient size and could enter the Canadian market. Thus the market could support only one indigenous CRS. Dissolution of the merger would therefore result in two inefficient Canadian CRS suppliers.
- Pegasus was a non-starter. It could not achieve an efficient scale on its own and, due to poor product design and Reservec's previously established position in the Canadian market, Pegasus could capture only low-volume, non-urban travel agencies.
- The economies of scale realized from the merger would allow the Canadian CRS system to invest in new technology and to provide high-quality service to smaller Canadian centres.
- Air Canada's and CAIL's desire to increase market share against North American rivals would keep Gemini CRS booking fees low for travel agents.

With respect to the impact on competition in the airline business the respondents stated that:

- the merger increased the ability of Air Canada and CAIL to negotiate improved reciprocal access to the North American CRSs, thus enhancing their overall access to the North American and international market; and
- CAIL benefitted from obtaining some of the “halo effect” Reservec formerly provided exclusively to Air Canada.

On the related issue of the merger facilitating future abuse of dominant position in both CRS and air travel, the respondents argued that:

- Reservec used to have 100 percent of the Canadian market but no anti-competitive conduct was detected by a federal task force, unlike the findings of the 1984 Civil Aeronautics Board with respect to the airline display practices of some U.S. systems;
- both Air Canada and CAIL were formally committed in writing to the Minister of Transport to allow fair access to Canadian CRSs;
- Gemini was operated autonomously from the airlines; and
- the DIR could make an abuse of dominance application at any time when the facts support it.

The Director’s application considered at some length Gemini’s provision of Last Seat Availability (LSA) only to Air Canada and CAIL and argued that this practice would lessen competition among Canadian airlines by creating an incentive to agents to rely on Air Canada’s and CAIL’s service exclusively. The respondents, however, contended that LSA had been overrated as an impediment to fair competition since travel agents could obtain the last seats directly from the airline in any event. They also contended that LSA provides an important commercial advantage in competing for travel agency clients with the “functionally superior” U.S. based Sabre system.

(b) Director’s Reply

The Director’s reply presented the following principal arguments:

(i) Market definition: Canada is the appropriate geographic market. CRS operations are geared to the principal business incentives of their airline

parents. U.S. CRSs, therefore, lack the same incentives as Gemini to provide service in Canada since their airlines cannot serve the entire Canadian market.

(ii) Critical mass/economies of scale: The respondents had presented no empirical evidence on the minimum efficient size for a CRS firm and no evidence that the Canadian market could support only one indigenous firm. Reservec was profitable before the merger with 2,900 travel agency subscribers.

(iii) Pegasus as a Failing Firm: Pegasus should be expected to lose money in its start-up years but by the time of the merger Pegasus had achieved 72 percent of the 1,000 travel agency subscribers originally forecast as required for break-even.

(iv) Remaining effective competition: Sabre's competitiveness in the Canadian market was entirely dependent on Gemini not exercising its market power by Gemini's refusal to provide direct access to airlines competing with Air Canada and CAIL.

(v) Abuse of dominance: Political assurances of fairness were insufficient protection to competitors. Wardair and other domestic competitors had no domestic CRS alternative as the result of the merger. Gemini had a near monopoly in some local CRS markets.

(c) Further developments

During the pre-hearing stages of the case it was announced (September 28, 1988) that the Gemini Group and the PARS reservation system based in the U.S. would become partners in a new CRS. This partnership would be the largest multi-hosted CRS in the world comprising the 3,500 Gemini and 6,300 PARS travel agency locations. The owner of PARS are Northwest Airlines and TransWorld Airlines.

On December 7, 1988, the DIR filed an amended application requesting the Competition Tribunal to order Air Canada and Canadian Airlines International Limited (CAIL) not to proceed with the PARS merger or, if that merger was complete, to order its dissolution.

The Director's application contended that the PARS merger, in addition to the Gemini merger, enabled Gemini and its airline owners to sell a share of their near monopoly in CRS in Canada to another company in exchange for the PARS CRS technology without diluting the market power that flows from the Gemini merger. The DIR contended that the PARS merger would not reduce either the incentive or the ability of Gemini to exercise that market power to prevent or lessen competition in Canada. Further, the DIR alleged that the PARS merger made it even more likely that Gemini would maintain or increase its "overwhelming market dominance" as follows:

As long as Air Canada and CAIL refuse to allow direct access links to other CRS systems from their completed inventorial schedule, fare rules, seat classes and seat availability, the enhanced functionality of PARS over Gemini makes it even more unlikely that any other CRS will be able to effectively compete in Canada. . . .

The PARS merger will also eliminate additional possibilities for enhanced CRS competition and entry in Canada. PARS could have sold its CRS software to either Reservac or Pegasus; become an independent entrant in Canada; or more likely become a joint venture partner with one of the Gemini airline owners with the other free to operate independently or join with one of the other major U.S. or European CRS's as a second major CRS competitor in Canada.

The DIR also alleged that the PARS merger supported his position that the Gemini merger was not necessary to obtain state-of-the-art CRS technology for Canada. Such technology would appear to be readily available in the U.S. and elsewhere, and Air Canada or CAIL individually could acquire such technology by a joint venture or by hardware or software purchases from a number of CRS vendors. Consequently, the DIR stated that Gemini did not need a "near monopoly" in Canada in order to build a "made in Canada" CRS and that indeed with the PARS merger it had chosen not to do so.

With respect to the impact of the PARS merger on competition in the airline industry, the DIR contended that Air Canada and CAIL would continue to have the incentive and ability to prevent or lessen competition in airline markets and that the PARS merger did not alter this at all. In addition, the DIR contended that the remaining partners in PARS, TWA and Northwest Airlines, would have no incentive to stop this lessening of competition within Canada because they did not serve domestic Canadian city pairs and would

not be adversely affected by any reduction in domestic airline competition. The DIR also noted that the U.S. Civil Aeronautics Board rules, requiring non-discriminatory CRS access, were not applied to the operation of PARS in Canada. As a result, PARS would be able to engage in CRS abuses in Canada without being subject to any CRS rules or regulations.

Finally, the Director's application contended that the current PARS data display was ordered in such a way that interlined connections were penalized and given a lower priority in the display as compared to on-line connections. No interline "penalty" would be imposed on connections between Air Canada/CAIL and their respective affiliates. The DIR contended that the inability of interlined carriers, such as Wardair, to have their flights and fares displayed without the substantial "penalty" imposed through the PARS merger prevented such carriers from competing with the two dominant Canadian carriers, Air Canada and CAIL. The application contended that the PARS "penalty" against interlined connections heightens the barrier to entry into Canadian city-pair markets by increasing the control of Air Canada and CAIL over feed traffic.

(d) Resolution

The Tribunal's hearing on the Director's expanded application was to have commenced on April 3, 1989. The Tribunal was critical of this long delay to trial and largely placed the blame on the tactics of legal counsel on both sides. On that date, the DIR advised the Tribunal that agreement had been reached with the respondents and that he would be seeking the Tribunal's approval of the consent.

The proposal in essence allowed the Gemini merger to remain intact but required Gemini to provide data links to competing CRSs. Following a hearing on the proposed Order, at which a competitor/intervener, American Airlines, provided both expert and factual evidence, the Tribunal, at a post-hearing conference indicated changes it would require to approve the Order.

The principal elements of the Order approved by the Tribunal on July 7, 1989 were:

- (1) Air Canada and CAIL must provide any other CRS operating in Canada with the same advance seat selection and boarding pass capability which they have provided to Gemini, if that other CRS offers reciprocity;

- (2) Gemini must make available to travel agents any and all enhancements made available to Gemini by participating carriers;
- (3) Gemini must provide direct access links to provide Last Seat Availability on Air Canada and CAIL, providing both “look but not book” and “look and book” features, subject to specified terms, implementation dates, and prices for certain existing competing CRS;
- (4) Booking fees must be non-discriminatory; and
- (5) The respondents would prepare a set of CRS rules affecting information display, contracts with participating carriers and subscribers (to prevent discrimination, tying or other exclusionary practices), access to airline information service enhancements, marketing information, ticketing and enforcement.

The reasons for its approval of the Order amplify the Tribunal’s approach to exercising its discretion to approve such orders. The key elements are:

- (1) The role of the Tribunal is not to ask whether the consent Order is the optimum solution to the anti-competitive effects which are assumed to arise from a merger;
- (2) The Tribunal’s role is to determine whether the consent Order meets a minimum test. The test is whether a merger, as conditioned by the terms of the consent Order, results in a situation where the substantial lessening of competition, which it is presumed will arise from the merger, has, in all likelihood, been eliminated; and
- (3) If the terms of an Order are vague and therefore cannot be enforced by way of contempt proceedings, or if the terms imposed are virtually impossible to monitor, then the Order cannot meet the test of effectiveness necessary to eliminate the substantial lessening of competition which is required of it.

In this context it would be worth setting out the concluding remarks of the Tribunal:

The determination of whether or not a given situation will result in a substantial lessening of competition is a speculative decision. An order such as that which the Tribunal is asked to issue is a web of

interrelated provisions. Counsel for the Director referred to it as a delicate balance of trade-offs. There is no doubt that there is more than one combination of terms and conditions which could achieve the result which it is hoped the terms and conditions which are now before the Tribunal will achieve.

There have been significant modifications made to the consent order in response to concerns raised during the course of the hearing of this application and in response to suggestions made by the Tribunal. A comparison of the consent order filed on April 13, 1989 and that filed on June 2, 1989 demonstrates this.

As noted above, the Tribunal has expressed concerns that have not been met. It may very well be that had the Tribunal crafted the order itself a set of conditions would have resulted different from those which the Director and the respondents have agreed upon. There is no doubt that if some of the provisions proposed by American Airlines had been adopted into the consent order a more rigorous instrument for creating a post-merger competitive environment would have been created. But, as has already been said, the Tribunal does not consider that it has been given a mandate to craft the best possible terms and conditions for protecting competition. Its role is limited to vetting the order before it to ensure that the proposed terms and conditions are likely to be effective in eliminating any adverse effects of the merger.

It is of considerable significance that almost all of the intervenors support the consent order, including American Airlines. It is of significance that there has been little evidence adduced that the merger as conditioned by the consent order will lead or will likely lead to a substantial lessening of competition. In addition, the Tribunal notes that the general trend is toward the formation of large, jointly-owned CRSs. It is clear that the implementation of some of the terms of the consent order will require the diligent and continual surveillance of the Director. It is clear that changed conditions or effective enforcement of the order may require a return to the Tribunal for either changes to or interpretations of the order. Taking all these considerations into account, the Tribunal concluded, on the basis of the evidence before it, that the consent order meets the test required by the legislation.⁹

E. ABUSE OF DOMINANT POSITION

1. Introduction

The 1986 revision established a new reviewable practice, abuse of dominant position, to replace the ineffective and unused criminal prohibition against creation or operation of a monopoly against the public interest.

Abuse of dominant position is broadly drafted, and could become potentially sweeping in scope, depending upon the priorities and creativity of the DIR and the initial jurisprudence of the Tribunal.

As discussed above, it is at least likely that abuse of dominant position will take the place of the predatory pricing offence as the means of reviewing and controlling unilateral anti-competitive pricing behaviour.

Unfortunately, it is not easy at this point to anticipate the likely scope or effectiveness of this reviewable practice. In the one decided case (NutraSweet discussed below) the Tribunal also granted substantially all the relief sought by the DIR through finding the existence of “lesser” reviewable practices (tied selling and exclusive dealing). There have been no enforcement guidelines published.

Abuse of dominant position is, however, modelled upon article 86 of the *Treaty of Rome* which established the general competition policy for intra-Community trade, and there is a considerable and growing body of European Community jurisprudence on the practice. In fact, the NutraSweet monopoly over aspartame was challenged under the *Treaty of Rome* prior to the DIR applying for a similar remedy in Canada.

For example, the market control concept of the *Treaty of Rome* relates to a position of economic strength which allows an undertaking to prevent effective competition being maintained in the relevant market. Such a position enables the company to behave to an appreciable extent independently of its competitors and customers, and the ultimate consumer. At first, this approach appears to be broader than the concept of market control advanced in North American anti-trust literature which (see the proposed merger guidelines) concentrates on the capacity to sustain price increases for a non-transitory period. In Europe, dominance has been found where the

supplier's overall market share was less than 50 percent but it had a much higher specific share by virtue of its brand recognition in certain product lines.

2. Statutory Provisions

Section 79 of the *Competition Act* authorizes the Tribunal to issue an Order prohibiting all or any respondents in a DIR's application from engaging in a practice of anti-competitive acts where:

- (1) the respondent(s) substantially or completely control a class or species of business throughout Canada or in any area of the country;
- (2) the respondent(s) have engaged in or are engaging in a practice of anti-competitive acts; and
- (3) the practice is having, has had, or is likely to have the effect of preventing or lessening competition substantially in a market.

If a remedial prohibition Order is insufficient to restore competition, the Tribunal has the jurisdiction to order divestitures of asset or shares sufficient to overcome the practice's effects.

In making its impact assessment, the Tribunal must address whether the practice is a result of superior competitive performance.

Section 78 of the Act is a non-exhaustive list of anti-competitive acts. These are:

- (a) squeezing, by a vertically integrated supplier, of the margin available to an unintegrated customer who competes with the supplier, for the purpose of impeding or preventing the customer's entry into, or expansion in, a market;
- (b) acquisition by a supplier of a customer who would otherwise be available to a competitor of the supplier, or acquisition by a customer of a supplier who would otherwise be available to a competitor of the customer, for the purpose of impeding or preventing the competitor's entry into, or eliminating the competitor from, a market;

- (c) freight equalization on the plant of a competitor for the purpose of impeding or preventing the competitor's entry into, or eliminating the competitor from, a market;
- (d) use of fighting brands introduced selectively on a temporary basis to discipline or eliminate a competitor;
- (e) pre-emption of scarce facilities or resources required by a competitor for the operation of a business, with the object of withholding the facilities or resources from a market;
- (f) buying up of products to prevent the erosion of existing price levels;
- (g) adoption of product specifications that are incompatible with products produced by any other person and are designed to prevent his entry into, or to eliminate him from, a market;
- (h) requiring or inducing a supplier to sell only or primarily to certain customers, or to refrain from selling to a competitor, with the object of preventing a competitor's entry into, or expansion in, a market; and
- (i) selling articles at a price lower than the acquisition cost for the purpose of disciplining or eliminating a competitor.

The common feature of these anti-competitive acts is an exclusionary purpose or effect. Thus, the presence or prospect of competitive entry or expansion into the relevant market (which need not be the market in which the respondent is dominant) would appear to be essential to achieve a remedy from the Tribunal.

Conversely, the practice would not appear to be directed to remedying monopoly practices by a firm in its dominant market where there was no real competition or no reasonable prospect of competition: in the extreme, a market supplied by a "natural" monopoly.

Apart from this qualification, the Act invites an "I know it when I see it" approach to identifying a practice of anti-competitive acts. In practical terms, these are defined by the nature of the way they lessen competition, the extent of market control of the dominant firm, and the extent of existing and prospective competition. Thus it is quite difficult to make specific or discrete advice on each element of the practice.

The Tribunal's remedial power, while broad, is cast as a prohibition aimed at removing the undesirable effect on competition. It therefore does not imply direct government regulation of the manifestation of dominance (for example, prices). And the Tribunal's discussion of enforceable merger consent Orders makes it clear that a properly fashioned dominance prohibition would have to permit the Tribunal to stand back and watch rather than participate with the dominant firm's management in restoring competition.

3. NutraSweet Case

(a) DIR Application

On June 1, 1990 the DIR filed a Notice of Application with the Competition Tribunal requesting Orders against the NutraSweet Company under the abuse of dominant position, exclusive dealing and tied selling provisions of the *Competition Act*.

NutraSweet is the principal supplier in Canada of aspartame, an intense sweetener used as a substitute for sugar in a variety of food products. The application noted that NutraSweet controls about 95 percent of the market for aspartame in Canada and that its principal competitor, Tosoh Canada Ltd., a subsidiary of Holland Sweetener Company of the Netherlands, supplied approximately three percent of the market.

The application contended that aspartame constitutes a distinct product market from sugar and other artificial sweeteners for a variety of reasons: it provides an alternative to sugar for diabetics and weight-conscious consumers; it is not harmful to the teeth; it is non-caloric unlike other bulk sweeteners; and it is distinctive from other intense sweeteners such as saccharin and cyclamates because of the much greater range of food products and applications for which it has been authorized under the Food and Drug Regulations.

The alleged anti-competitive acts, exclusive dealing and tied selling requirements, flow from the same set of alleged contractual provisions which, as a whole, the DIR contend had prevented the creation of a separate Canadian market for the sale of aspartame. The DIR's application alleged that these acts constituted predation and sophisticated price discrimination practices.

The acts alleged by the DIR included:

- (1) Use of NutraSweet's U.S. position to foreclose competition in other countries through entering into worldwide exclusive supply contracts of several years' duration with the parent companies of the largest purchasers of aspartame in Canada (principally soft drink manufacturers).
- (2) Inducing exclusivity through the structure of its supply contracts. The relevant provisions include customer obligations to purchase their entire supply of aspartame from NutraSweet; customer obligations to use aspartame as the sole sweetener in their products; and a variety of fidelity rebates which the DIR alleged are designed to induce customers to purchase all of their aspartame requirements from NutraSweet. These fidelity rebates were alleged to include volume discounts, incentives for encouraging others to purchase aspartame from NutraSweet, and allowances for the display of NutraSweet's trademark or brand name on a customer's packages. The DIR contended that, in order to remain competitive, a NutraSweet customer must take advantage of all the rebates offered by NutraSweet under its supply contracts. As a result, the customer was effectively forced to purchase its entire supply of aspartame from NutraSweet and to affix NutraSweet's trademark to its packaging.
- (3) Extending patent rights through exclusive contracts. The DIR contended that, immediately before the expiration of its Canadian patent, NutraSweet negotiated a number of long-term, exclusive supply contracts for aspartame which, the DIR alleged, ensured the maintenance of NutraSweet's dominant position in the Canadian market regardless of whether its patent had expired.
- (4) Creation of market transparency to control competition. The DIR alleged that, through contractual terms with its customers (which were ostensibly aimed at maintaining the customers' competitiveness), NutraSweet was able to obtain knowledge of its competitors' activities and to meet the competition in a fashion that substantially lessened competition. The DIR contended that these agreements take several forms: "Meet or release" clauses (also known as "English clauses") which allow customers to obtain competitive offers for aspartame but also permit NutraSweet to meet the competitor's price if it were lower or to release the customer from the supply contract if NutraSweet elected not to meet the offer; an "extended release" clause which permits one customer to go to

another for aspartame in the event that another third-party customer of NutraSweet aspartame is released from its supply contract; a “most favoured nation” clause which requires NutraSweet to ensure that the customer is not charged a price for aspartame which places it at a competitive disadvantage in its own industry.

- (5) The combination of most favoured nation and fidelity clauses. The DIR alleged that the combination of these terms induce customers to purchase their entire supply of aspartame from NutraSweet because customers can become free riders on the lower prices obtained by other aspartame customers and therefore have no incentive to seek lower prices on their own.
- (6) Abuse of trademark. The DIR contended that NutraSweet’s practice of providing allowances to those customers which place NutraSweet’s trademarks on their products creates barriers to entry since the competitor’s price must both justify the cost of removing the trademark and cover the foregone trademark display allowance.
- (7) Selling below acquisition cost or long-run average cost. The DIR also alleged that the net prices charged to certain Canadian customers after all of the above discounts, allowances and rebates are less than NutraSweet’s average acquisition cost or long-run average cost in production of aspartame.

Accordingly, the DIR contended that these alleged anti-competitive acts have foreclosed the aspartame market in Canada to potential alternative suppliers; the Canadian market is not open to competition at all due to the worldwide contracts negotiated outside Canada which require exclusive use of NutraSweet’s aspartame in Canada; and there is no incentive for customers to seek alternative sources of supply because of the price preferences and other measures noted above.

The DIR also relied on the contractual provisions which he alleged constitute anti-competitive acts as the basis for asserting that NutraSweet has practiced both exclusive dealing and tied selling.

To establish exclusive dealing, the Director’s application referred to the alleged presence of exclusive worldwide supply contracts; exclusive supply provisions in individual contracts; the affixing of the NutraSweet trademark on customers’ product labels; exclusive dealing inducements through the

presence of English clauses; most favoured nation clauses; trademark display allowance provisions; cooperative marketing rebate schemes; fidelity rebates; and free product clauses in NutraSweet's arrangement with its customers.

The DIR alleged that tied selling had occurred because NutraSweet, as conditions for the supply of NutraSweet's aspartame required customers to affix the NutraSweet trademark to their products and to refrain from using another brand of aspartame in conjunction with NutraSweet's brand. The DIR also alleged that the provision of substantial allowances for trademark display in a number of NutraSweet's supply contracts together with other fidelity clauses constituted inducements to meet these conditions.

The Orders requested by the DIR relating to abuse of dominant position, exclusive dealing and tied selling were similar. They prohibited world-wide contracts with multinational customers governing the supply of aspartame to Canadian affiliates; they prohibited the requiring of customers to purchase their entire supply of aspartame from NutraSweet; they prohibited the selling of aspartame to Canadian customers at below acquisition cost, and the granting of customer price concessions and other allowances not available to competitors of the customer for the same agreed volume of aspartame.

As well, the DIR applied for a declaration that the contractual provisions noted above were of no force or effect and could not be judicially enforced by NutraSweet. The DIR also requested the Competition Tribunal to order NutraSweet to include a most-favoured-nation clause in all supply contracts of customers which compete with each other in selling products containing aspartame if such a clause was included in a contract with any of them.

Finally, the DIR sought a declaration that any contracts entered into by NutraSweet or its affiliates pursuant to worldwide contracts were of no force or effect insofar as they affected the supply of aspartame in Canada and that NutraSweet could not require minimum exclusive annual volume commitment in any supply contract which was greater than 50 percent of a customer's total annual volume requirements of aspartame.

(b) NutraSweet's Reply

The NutraSweet response, filed on July 25, 1989, contended that the firm's success in the market was purely the result of innovation, risk-taking and superior competitive performance. The respondent contended that its superior competitive performance was the result of continuous research and development, testing of aspartame to obtain approvals in various countries, intensive promotion, and a superior product and supply network, none of which had been undertaken by its competitors.

NutraSweet also contended that it had no market power in the supply of aspartame for a number of reasons including the following:

- (1) extensive in-house research and development of artificial sweetener is now being conducted by manufacturers of food products;
- (2) large food and beverage manufacturers can readily manufacture aspartame once it comes off patent;
- (3) competition and supply of aspartame will intensify towards the end of NutraSweet's U.S. patent in 1992;
- (4) no barriers to entry exist in Canada for manufacturers that currently supply aspartame.

NutraSweet also noted that its principal customers were large and sophisticated and could readily choose among competing suppliers; that prices for aspartame have been declining rather than increasing; and that contracts for the supply of aspartame were freely negotiated and had short time commitments that facilitated frequent competitive overtures from other suppliers.

NutraSweet denied that aspartame might be regarded as a distinct product market. The firm suggested that aspartame, as well as many other natural and artificial sweeteners, competed for the same ultimate customers and that aspartame fulfilled the same purpose as all other sweeteners. NutraSweet also noted extensive research and development activity by drug and food companies to improve and invent artificial sweeteners.

With respect to the geographic market, NutraSweet contended that Canada was not the appropriate geographic market, but that, rather, given the

very low cost of shipping artificial sweeteners and the number of plants in existence in a variety of countries, the geographic market was global.

NutraSweet contended that the contractual practices which the DIR regarded as evidence of policy to maintain market pre-eminence were, in fact, all provisions sought by NutraSweet's customers to secure benefits for themselves or were normal commercial practices.

NutraSweet claimed that exclusive use and worldwide contracts were negotiated by customers to obtain secure supplies and consistent high-product performance. Volume discounts and cooperative marketing are regarded as normal commercial practices by NutraSweet. Trademark display allowances provide a benefit to customers through cost reductions of aspartame and, in turn, facilitate NutraSweet in protecting the goodwill of its trademark. "Meet or release" and "price protection" clauses, NutraSweet contended were actually sought by its customers to permit them to maintain their own competitiveness, particularly in the soft drink market.

(c) Tribunal Decision

The Tribunal's decision was published on October 4, 1990. The Tribunal concluded that the practices of abuse of dominance, tied selling and exclusive dealing were present and issued a remedial Order.

At the time of writing, NutraSweet had appealed the decision to the Federal Court, and the DIR had cross-appealed seeking different remedies than those imposed by the Tribunal.

With respect to the issue of market definition, the Tribunal examined the market for all sweeteners and assessed the cross-elasticity of demand between high-intensity and high-calorie sweeteners and regulatory barriers facing suppliers of high-intensity sweeteners. The Tribunal found that there was no direct evidence of competition between aspartame and caloric sweeteners, and only weak evidence of indirect competition between high-intensity sweeteners. However, the Tribunal felt that no other high-intensity sweetener was a good substitute for aspartame across large market segments. The Tribunal therefore decided to define the product market as aspartame since it mattered little whether the definition was "aspartame" or "high-intensity sweeteners" because of the very limited degree to which non-aspartame, high-intensity sweeteners were present in the Canadian market.

In assuring market control, the Tribunal focussed on entry barriers. The Tribunal concluded there were significant entry barriers comprising patent portfolios, economies of scale and sunk costs.

The DIR had argued that market control meant control over supply as defined by the normal dictionary meaning. NutraSweet argued that it meant “market power.” Market power would mean an ability to set prices above competitive levels for a considerable period.

The Tribunal accepted the respondent’s position, stating however:

This finding is of little practical import because, ultimately, all relevant indicators of market power must be considered in determining whether there is likely to be a prevention or lessening of competition substantially.

Having accepted the respondent’s definition of control, the Tribunal proceeded to assess whether large buyers, such as Coca-Cola and Pepsi, were able to protect their interests through their own strong bargaining position. The Tribunal concluded that, although Coca-Cola and Pepsi have considerable resources to protect their interests, this did not in and of itself eliminate NutraSweet’s market power. The Tribunal concluded:

The evidence that NSC possesses appreciable market power given its market share (over 95% of sales in Canada), entry conditions and the constraints operating on its largest customers is sufficiently compelling so that the boundaries of substantial control need not be explored. Its “control” is clearly substantial.

With respect to “class or species of business” the DIR argued that these words should be interpreted in the commercial sense rather than in an economic sense, and that, accordingly, the business of NutraSweet was the manufacture and supply of aspartame. The Tribunal took the view of the respondent, however, that “class or species of business” is the same as the relevant product market.

The Tribunal’s interpretation of “class or species of business” is an important decision, particularly since it contradicts the Eddy Match decision of the

Quebec court, under the old monopoly section. The Tribunal justified this distinction in the following manner:

Based on the facts in *Eddy Match Co.* and the different legislative schemes of the *Combines Investigation Act* and the *Competition Act*, the Tribunal does not believe that this case provides a sound basis for identifying “class or species of business” without referring to possible substitutes. The court in that case was seized with charges under a criminal statute, a case in which the accused had engaged in highly aggressive conduct towards other producers of wooden matches; Eddy certainly acted as though wooden matches were sufficiently distinct so that it was worthwhile for it to concentrate its efforts on that industry. In the present statute, however, s.79 provides other remedies and the deciding body is a specialized tribunal. It would run contrary to the spirit of this legislation for the Tribunal to eschew other relevant factors (that is, possible substitutes) on some presumed technical ground.

The decision by the Tribunal that it will not necessarily follow previous jurisprudence under the *Combines Investigation Act* in criminal matters is an important indication of the Tribunal’s view of the purpose of the 1986 revisions.

The Tribunal then considered the requirements of a “practice of anti-competitive acts” as contained in section 79. In this area also, the Tribunal’s decision is an important precedent. It stated:

This list of anti-competitive acts is clearly not meant to be exhaustive and the respondent admits that other conduct not specifically mentioned in s.78 can constitute an anti-competitive act. A number of the acts share common features but, as recognized by the Director and the respondent, only one feature is common to all: an anti-competitive act must be performed for a purpose, and evidence of this purpose is a necessary ingredient. The purpose common to all acts, save that found in paragraph 78(f), has an intended negative effect on a competitor that is predatory, exclusionary or disciplinary.

This finding indicates the Tribunal believes there must some element of intent or purpose to constitute an anti-competitive act.

The Tribunal interpreted “practice” as follows:

If there is a good reason to avoid a limiting interpretation of “practice” under criminal law, it is all the more important to do so under s.79. The anti-competitive acts covered in s.78 run a wide gamut. Some almost certainly entail a course of conduct over a period of time, such as freight equalization in paragraph 78(c), whereas others consist of discrete acts, such as the setting of product specifications in paragraph 78(g). The interpretation of “practice” must be sufficiently broad so as to allow for a wide variety of anti-competitive acts. Accordingly, the Tribunal is of the view that a practice may exist where there is more than an “isolated act or acts.” For the same reasons, the Tribunal is also of the view that different individual anti-competitive acts taken together may constitute a practice. It is important to stress, however, that this does not in any way relieve the Director of the burden of establishing an anti-competitive purpose for each of the acts.

The Tribunal then considered how one would establish an anti-competitive purpose. The DIR submitted that this test could be met through evidence of subjective intent, such as verbal or written statements of personnel of the respondent, or through a consideration of the nature of the act itself, the latter being based on the premise that a corporation intends the necessary and foreseeable consequences of its acts. The Tribunal accepted the Director’s position. It also indicated that, in most instances, the purpose of a particular act will have to be inferred from the surrounding circumstances.

With respect to meet or release clauses, the Tribunal found that large customers, such as Coca-Cola or Pepsi, used them as a way of mitigating the effect of being locked into an exclusive contract. The Tribunal stated:

If exclusive supply is objectionable in the instant case, so is a meet or release clause: by making exclusivity more acceptable to customers it serves as an inducement for customers to enter into exclusive arrangements.

With respect to the most favoured nation clause whereby NutraSweet contractually agreed to offer its best price in its contracts, the DIR had argued that this was an inducement to exclusive dealing. The Tribunal found that:

The Director submits, correctly in the Tribunal's view, that only a firm with a very large market share can be expected by its customers to provide a most-favoured-nation clause because only it will almost certainly be selling to the customers' competitors.

However, the Tribunal appeared not to find all of these practices as anti-competitive acts. The Tribunal stated:

The Tribunal sees little purpose in the context of the present case in determining whether each clause constitutes an anti-competitive act. It is doubtful whether the meet or release and most-favoured-nation clauses would exist in the absence of an explicit or implicit exclusive supply agreement. In the Tribunal's view, the issue is whether the agreements requiring exclusive supply and all the contract terms related to it, have an exclusionary purpose. The Tribunal is persuaded that this is the case.

The DIR had alleged that NutraSweet had engaged in the anti-competitive act set out in paragraph 78(i) of "selling articles at a price lower than the acquisition cost for the purpose of disciplining or eliminating a competitor." The Tribunal accepted the respondent's argument that acquisition cost is not easily applied in a manufacturing situation. The Tribunal also found that the language of the paragraph suggested that Parliament intended it to be applied mainly to distribution situations where the articles would be resold.

While limiting the application of paragraph 78(i), the Tribunal considered the broader issue of whether there could be an anti-competitive act or some form of predatory pricing not specified in section 78 but nevertheless, caught by the dominance provisions. The Tribunal, in an important conclusion, stated that it was satisfied that the section was broad enough to cover other forms of predatory pricing. The Tribunal then considered the appropriate test for predatory pricing and basically concluded that the Areeda-Turner test is the appropriate standard (that is, prices are presumed not to be predatory if they exceed average variable costs). This is similar in principle to the Stage II test under the Bureau's proposed Predatory Pricing Bulletin.

The DIR had alleged that NutraSweet was using its patent in the United States to gain a competitive advantage in Canada. This U.S. market power was alleged to have been used with Coca-Cola and Pepsi in particular. The

Tribunal accepted the Director's position that the use of a monopoly position in one country to obtain a competitive advantage by a dominant firm in another market constitutes an anti-competitive act.

With respect to proof of substantial lessening of competition, the Tribunal stated:

The factors to be considered in deciding whether competition has been or is likely to be substantially lessened are similar to those that were discussed in concluding that NSC has market power. In essence, the question to be decided is whether the anti-competitive acts engaged in by NSC preserve or add to NSC's market power.

In summary, the particular acts the Tribunal found to be anti-competitive were exclusive supply and use clauses, logo display allowances, cooperative marketing allowances, meet or release clauses, and most-favoured-nation clauses. The Tribunal also included the use of U.S. patents to foreclose competition by a system of rebates on exports from the United States to induce Canadian importers to have only NutraSweet aspartame in their products. The Tribunal, therefore, concluded that those practices were having the effect of preventing or lessening competition substantially.

IV. COMPETITION ACT COMPARED WITH NATIONAL TRANSPORTATION ACT, 1987 AND RELATED LEGISLATION

A. INTRODUCTION — HISTORICAL PERSPECTIVE

The *National Transportation Act, 1987* (NTA, 1987) represents the most recent legislative milestone in a process of staged reduction in direct federal regulation of intercity transportation services that has continued since the 1960s.

However, federal regulatory policy for the transportation industry was not static between the 1968 and 1987 revisions to this industry-specific legislation. For example, during this period, largely due to competitive pressures from expansions in trucking as a substitute to rail freight and from passenger air deregulation in other jurisdictions, regulatory and policy changes increased the operating flexibility of rail service and air carrier managers.

To some extent, particularly in relation to air carrier regulation, the 1987 revision mirrored rather than advanced prevailing regulatory policy of the Canadian Transport Commission.

However, the NTA, 1987 did not significantly alter the principal institutional and administrative elements established in the NTA of 1967, for the delivery of the federal government regulatory product for this sector.

By way of contrast, as discussed in Parts II and III, the 1986 revision of the *Competition Act* included both major substantive changes geared to increase effectiveness and predictability as well as major institutional changes involving the establishment of the Competition Tribunal and “decriminalization” of Canada’s merger and monopoly laws of general application. These changes have facilitated a more hands-on and proactive approach to competition law enforcement.

In fact, the substitution of civil for criminal process and substantive standards can be regarded as an overall trend in the evolution of Canadian competition law that is not confined to the 1986 revision. Prior to the 1976 revision of the *Combines Act*, the law was entirely based upon criminal law prohibitions. The 1976 revisions introduced a range of reviewable practices relating to non-price vertical restraints to competition which were regarded as manifestations of inefficient monopolization — as well as new criminal prohibitions respecting unfair and inefficient marketing practices. These prohibitions are now being reviewed by the Bureau as possible candidates for decriminalization. Charter requirements may ultimately force decriminalization of the conspiracy prohibitions. And, in light of the structure of the abuse of dominant position practice, it is now probable that this provision, together with the Bureau’s draft predatory pricing policies, and not the criminal predatory pricing prohibitions, will be the general future context for Bureau action respecting anti-competitive unilateral pricing practices.

This comparison of the *National Transportation Act, 1987* and the *Competition Act* looks at both the institutional arrangements for law enforcement and the substantive law established by each Act.

The comparison is based on a point-in-time snapshot of two streams of economic regulation. It is probable, in my view, that enforcement and substantive provisions of both Acts will be significantly amended within the decade.

B. NATIONAL TRANSPORTATION ACT, 1987 AND RELATED LEGISLATION

As noted in Part II, the *Competition Act* does not apply, by operation of common law, to commercial activities that are otherwise authorized or regulated by valid legislation. Nor does it apply to agents of the Crown. The 1986 revisions have, however, made the *Competition Act* applicable to certain Crown corporation activities. And, as noted below, this exemption of regulated conduct can be, and has been, overridden by legislation.

1. Substantive Provisions

The *National Transportation Act, 1987* and related legislation establishes discretionary regulatory authority to control the following intercity commercial behaviour of intercity passenger transportation service suppliers:

- *Intercity rail transport:*

- (a) entry is dependent upon a public convenience and necessity finding;
- (b) tariffs are subject to review and variance if found not to be in the public interest. (By contrast, rail freight rates are required to be “compensatory” using a standard based on a statutory average variable cost coverage that has been elaborated by extensive regulatory cost accounting rules.)

- *Intercity domestic air transport:*

- (a) Southern and Northern Canada (“designated area”):

Domestic service basic fare increases may be rolled back or disallowed, upon complaint, if unreasonable and if there is no alternative, effective, adequate and competitive transportation service.

- (b) Northern Canada (only):

Limited entry and exit restraint based upon locality and assessment of impact on competitors.

Service level and points served subject to licence restrictions.

- *Intercity bus transport:*

Jurisdiction delegated to individual provinces which employ a range of discretionary entry, exit and price controls of varying intensity and effectiveness. No material substantive changes have been made in the *Motor*

Vehicle Transport Act (MVTA), 1987 compared to the MVTA of 1954 with respect to extra-provincial bus transportation. Provincial regulatory policies all involve an element of protection of incumbent firms from competitive entry to help sustain a higher level of service (and lower prices) in less dense markets. In some cases, overall limits to firm profitability have been applied through the tariff approval process. Corporate profitability constraint appears to have been of diminishing importance over the last decade.

The NTA, 1987 also retains certain *non-discretionary regulatory controls*:

- (a) Acquisitions of Canadian Transportation Undertakings (all intercity passenger rail, bus and air suppliers) having in excess of \$10 million in assets or annual sales are subject to prior public notice and approval (which approval may be deemed after a certain period of time by operation of the statute). Disallowance requires a finding that the acquisition is against the public interest.
- (b) Entry by non-resident air carriers is subject to compliance with state-to-state agreements which comprehend carrier designation, points served, freedom of the air and capacity offered.
- (c) Non-resident voting share investment in domestic air carriers is limited to 25 percent.

The NTA, 1987 expressly provides that its regulatory regimes affecting base air fares and acquisitions do not affect the operation of other Acts of Parliament. Accordingly, the *Competition Act's* merger and abuse of dominance jurisdiction (to the extent that high pricing may be construed as an anti-competitive act) continues to operate and to run in parallel to these NTA regimes.

2. Institutional Features

The *principal institutional features* of the 1987 NTA revision are:

- Implementation and enforcement of discretionary and non-discretionary regulatory controls by an independent expert tribunal, the National Transportation Agency (the Agency), supported by its own staff and having an option to exercise its substantive powers following public hearings, and possessing disclosure powers to support its enquiries. The

structure of this regulatory agency is essentially the same as its predecessor the Canadian Transport Commission (CTC). The Agency and CTC are therefore not distinguished in our analysis of the behaviour of regulatory agencies (section D below). *The absence of a dynamic restructuring of the regulatory agency at the same time that federal regulatory discretion over the transport sector was being reduced, in my view, has created a potentially unhealthy tension between the new regulatory law and the institutional values and methods of the Agency that applies it.*

- Expanded opportunities for Cabinet intervention in Agency decision-making through binding policy directions to complement the pre-existing power to vary or rescind any Agency decision. (This power has, to date, proven to be academic).
- Reformulation of an elaborate "Parliamentary statement of national transportation policy" which the Agency is expected to implement through the exercise of its regulatory discretion. This policy seeks a safe, economic, efficient and adequate network of viable and effective transportation services making the best use of all available modes of transportation at the lowest total cost. As implementation guidance, the policy identifies:
 - (i) competition and market forces as the prime agents;
 - (ii) economic regulation only where necessary to serve the transportation needs of shippers and travellers but with no "unfair" limit on competition;
 - (iii) balancing commercial viability with regional economic development objectives;
 - (iv) that carriers bear a "fair share" of the costs they cause;
 - (v) that carriers receive "fair" compensation for public obligations; and
 - (vi) that fares be non-discriminatory and not create an undue obstacle to trade.

This policy statement arguably places more emphasis on the role of competition and market forces than did its predecessor in the NTA of 1967.

However, the potential impact of this shift, or any one of the enumerated policy goals for that matter, is left uncertain by the sheer variety of policy statements and their contradictory implications. Parliament, in my view,

has through unclear legislative policy (particularly when compared to the *Competition Act*) still left the Agency with ample room to act as an arbitrator of competing political and economic interests and to perpetuate institutionalized economic performance values carried over from the CTC era.

C. JURISDICTIONAL RELATIONSHIPS

1. Retail Level

(a) Exclusive Transportation Regulation Jurisdiction

Based on the considerations presented in Part II, at present, federal transportation legislation, and not the *Competition Act*, would apply only to the following dimensions of intercity bus and rail service:

- unilateral high or low pricing behaviour or discriminatory tariff terms of intercity rail service suppliers;
- discriminatory or exclusionary conduct arising from the geographical scope and level of VIA rail services; and
- in provincial jurisdictions where intercity bus tariffs require regulatory approval, both unilateral high and unilateral low pricing behaviour by such suppliers, as well as the terms and conditions of service provision.

The *Competition Act's* prohibitions against predatory pricing and abuse of dominant position (where the grounds are price gouging) would not apply in these circumstances.

Clearly, the remaining direct regulatory measures at the retail level which are applicable to intercity passenger transportation services have become quite limited. It would also be fair to say that the rate of erosion of direct economic regulation of retail pricing and terms of service has varied among modes.

Common carrier regulation for intercity bus service by blanket delegation of federal jurisdiction remains in the hands of provincial governments. In terms of process and substantive decision rules, this form of regulation remains largely the same as it has been for 20 years.

Passenger rail regulation jurisdiction is largely unchanged although there is increased scope to regulate service access for persons with disabilities and greater Cabinet and corporate discretion over the geographical scope and level of service.

Retail pricing regulation in intercity air transport is now restricted to a discretionary power to roll back basic service price increases. To date this power has not been exercised. Low pricing, and scheduled service terms and conditions and service levels (outside the designated area, that is, Northern Canada) are now not subject to direct regulation.

(b) Parallel Jurisdiction

Two areas of parallel jurisdiction confirmed by statute have already been noted: (1) high unilateral pricing by air carriers within Canada; and (2) mergers/acquisitions among transportation service suppliers.

However, there are major differences in the manner in which these jurisdictions may be exercised under federal regulatory and competition law.

These differences are examined below with respect to procedural administrative matters.

There is also, in my view, parallel regulatory and *Competition Act* jurisdiction in relation to certain structures or conduct of intercity passenger service.

First, the conspiracy provisions of the *Competition Act* may be employed to challenge high or low pricing behaviour of the intercity bus industry, even where this conduct is approved individually for each firm under a prior provincial tariff approval requirement, where such behaviour flows from an agreement among suppliers. This position is based upon a more restrictive interpretation of the Law Society of British Columbia case (see Part II, section B) than advanced by some commentators. Specifically, it is my view that more recent jurisprudence has confirmed that the conspiracy prohibition would be ousted in relation to the conduct of regulated firms or professions *only* where the agreement is authorized either expressly or by necessary implication by valid enabling legislation.

It should be kept in mind that the governing structure of the legal profession considered in the Law Society of British Columbia case is a statutorily

required collective agreement regime. However, provincial bus regulation is generally premised upon individual tariff applications being made by individual licensees rather than upon collective rate approval. Conceivably, there could be an exception from the conspiracy provisions where the provincial bus regulation regime expressly provides for the filing of joint tariffs by suppliers in the same geographic market (as opposed to joint tariffs respecting the provision of a particular service such as transportation between A and B using more than one licensee).

On the same basis, the conspiracy provisions of the *Competition Act* would have parallel application with the NTA, 1987 with respect to air carrier entry, exit and terms of service matters (to the extent the latter is not prescribed by regulation but is rather determined by the Agency on a case-by-case basis) which remain subject to some direct Agency regulation in the designated area.

Further, it is arguable that the definition of abuse of dominant position is sufficiently broad and flexible to apply to some conduct of an intercity bus or air carrier (in the designated area) where:

- the carrier's market dominance is the result of restrictive regulatory entry policies; and
- the regulator either lacks the jurisdiction or the practical opportunity to order cessation of the anti-competitive acts or to eliminate the opportunity to conduct them through either franchise elimination (delicensing) or allowing greater competition by licensing alternative suppliers.

As a practical matter, however, the Bureau of Competition Policy would probably not assert jurisdiction at this time in such a situation due to (1) the lack of jurisdictional certainty, (2) the likelihood that the markets involved are intra-provincial as opposed to national, and (3) the minor impact on economic efficiency and consumer welfare that the anti-competitive acts would entail.

(c) Exclusive Competition Act Jurisdiction

At the retail level, therefore, in my view, the *Competition Act* is the exclusive trade practices regulation instrument with respect to:

- predatory pricing and abuse of dominant position by air carriers (that is, pricing and other acts aimed at eliminating competition);

- resale price maintenance by air, rail and bus services; and
- non-price vertical restraints such as refusal to deal or tied selling practices by all modes of intercity passenger services where the transportation service is the tying product.

Because of the mode-specific nature of existing transportation regulation, it appears to be fairly clear that the *Competition Act* would apply exclusively to agreements to lessen competition (for example, price fixing and market sharing), including bid-rigging, involving more than one mode and non-price vertical restraints (for example, fixed selling and market restriction) between modes.

(d) Remaining Uncertainties

As discussed in Part III, section B, “article” under the *Competition Act* includes evidence of the right to transportation. The prohibition of price discrimination in this Act applies to discriminatory sales among competitors and also to discriminatory promotional allowances. Where the retail price of an intercity transportation service requires regulatory approval, it is arguable that the price discrimination prohibition does not apply at common law, since any discrimination has been lawfully authorized. However, where no express price approval is required, but there is a complaint-based appeal mechanism, it is arguable that there is room for application of the price discrimination prohibition given its highly specific nature. In either case, it would appear that the prohibition against discriminatory promotional rebates would apply to the passenger service supplier since the focus of the *Competition Act* is on a collateral service rather than on the regulated transportation service itself.

2. Input Market

(a) Introduction

Apart from being mode-specific, perhaps the principal jurisdictional difference between transportation regulation laws and the *Competition Act* is that the former are focussed upon the retail level of trade while the *Competition Act* applies to both the input and retail levels of an industry.

Transportation regulation laws are to a large extent responses to perceptions either that the retailer of passenger services would have excessive

market power without regulation, or alternatively, that market power should be generated by regulatory protection and directed to facilitate the pursuit of social policy objectives. *Such laws therefore do not (apart from the limited area of railroad access) provide a comprehensive instrument for eliminating market failures and inefficiencies that increase ultimate consumer costs or reduce ultimate consumer choice. Stated another way, transportation regulation laws do not remedy a market failure or excessive market power relating to inputs to the regulated service.* At the same time, they represent a legislative assumption that the most serious market failure or market power problems exist at the retail level.

Such an assumption may be initially valid but it is difficult to maintain as market conditions evolve. A comparative virtue of the *Competition Act* is that the legislation itself makes no assumptions about where in the production process the most serious forms of market failure or excess market power are likely to be found. Thus, subject to the reservations discussed in (c) below respecting markets subject to very high entry barriers, competition law is arguably better positioned to attack the root causes of consumer welfare loss (the disease) as opposed to the end result of these causes (the symptoms).

By way of example, the Bureau of Competition Policy, and not the Agency or federal Transport Department, was the source of a remedy to preclude exclusionary practices from developing in the dominant Canadian computer reservation system (CRS) firm. There is no longer any industry, travel agent or consumer support to legislate common carriage requirements based on snapshot observations of what may or may not be a "natural monopoly." Since the Competition Tribunal's resolution of the Gemini/Reservec case (see Part III, section D), there do not appear to have been substantial complaints of preferential display or access from air carrier competitors, travel agency consumers, or suppliers of collateral services (for example, ground transportation and accommodation).

(b) Competition Act Jurisdiction

The starting proposition therefore is that the *Competition Act* has exclusive jurisdiction in relation to passenger transportation service inputs. However, this jurisdiction is subject to two very basic limitations. First, for conduct to be actionable there must be a business relationship and also generally priced transactions. The evidentiary requirements of the Act's criminal

prohibitions and reviewable practices cannot for all practical purposes be met where the supply of goods or services is not being rationed by private sector markets using a price system. This effectively removes the Act from the supply of public goods or government services whether as inputs or as end products.

The supply of government services (whether they are priced as otherwise supplied under a contract, or not) is also effectively exempted by the common law exemption relating to Crown Agents. As noted, the exemption was cut back in the 1986 revisions to the Act only with respect to the commercial activities of Crown corporations.

Thus, to the extent that market failure or excessive market power in government-supplied inputs to passenger transportation services have a negative impact on consumer welfare, such problems remain out of reach of the *Competition Act* and are subject to correction, if at all, only through Ministerial accountability. It is arguable that this accountability structure is least susceptible to the introduction of the market-creation measures implicitly favoured by competition laws to remedy problems of inefficient resource allocation.

(c) Comparison of Substantive Tests

Regulation legislation for federal and provincial passenger service has largely adapted “public interest” and “reasonableness” tests to the determination of whether regulatory intervention is required in price, merger and supply decisions. Certain provincial statutes for the regulation of intercity bus service market entry and the *Railway Act*, with respect to railway market entry, contain a “public convenience and necessity” test.

All three tests are in practice equally broad. In fact, as legislative articulations of government decision-making discretion go, they are probably the broadest legislative grants of discretion that can be made. In the case of NTA, 1987 powers, it is possible to argue that the Act’s statement of transportation policy limits regulatory discretion. However, in reality the policy statement is so extensively qualified and rife with contradictory objectives that can only be balanced on a case-by-case basis, that no real limitation to regulatory discretion has been created.

Thus, these federal and provincial regulatory laws effectively permit the decision maker to apply any set of economic or social criteria to define appropriate regulated firm conduct. Stated another way, there is no externally imposed reference for what is, or is not, in the consumer interest. Identification of that becomes a self-referential exercise based on references to and analysis of the regulatory jurisprudence under each regime together with the balance of interests represented in the overall decision-making process.

Accordingly, there is nothing in the substantive tests of existing transportation regulation laws that would prevent the decision-makers from adopting the standards of anti-competitive behaviour of the *Competition Act* as, at the very least, difficult-to-rebut presumptions of behaviour that is not in the consumer interest even though administrative law rules require the decision making to retain some discretion and objectivity to support case-by-case adjudication.

On the other hand, the decision-making tests of the *Competition Act*, reviewed in Part III, are much more highly articulated and clearly focussed upon economic concepts and standards. This Act itself provides considerable guidance in such areas as the definition of anti-competitive acts of a dominant firm, the characteristics of anti-competitive non-price vertical restraints to trade, and the criteria for merger impact assessment.

Additional detailed guidance has been provided by the Bureau itself on key matters such as market definition of what is unreasonably low pricing, and what constitutes a substantial lessening of competition. The level of detail in this guidance may appear daunting but it is designed to permit private interests from anticipating Bureau action with the greatest possible precision. The Bureau is in a better position to provide such guidance and to bind itself to it as its role is administrative and does not involve adjudication. The Bureau is therefore not constrained by administrative law rules that require regulatory agencies to retain impartiality and decision-making discretion.

In my view, the criteria of the *Competition Act*, as supplemented by the DIR guidelines on price discrimination, predatory pricing and mergers discussed in Part III, are to a high degree consistent with prevailing microeconomics theory in the situations where individual firm behaviour is likely to result in a net reduction of consumer welfare within the economy as a whole.

That is not to say that firm behaviour which is not actionable under the *Competition Act* would never disadvantage particular consumers or groups of consumers for a certain period of time. Such a guarantee of course cannot be made in a competitive market system where government trade regulation is designed to be applied on an exceptions basis. Rather, the *Competition Act* addresses the overall impact on consumer welfare in the product and geographic markets served by particular firms. Through the use of economy-wide conduct standards, the Act attempts to ensure that all such markets are subject to the same restraints and therefore that all markets operate under the same incentives and opportunities to maximize outputs and profits.

The basic assumptions of competition law are that:

- (1) markets are subject to competition; and
- (2) markets (including both buyers and sellers) optimally respond to prevailing price signals.

The first assumption, however, cannot be readily satisfied if the industry is most efficiently supplied by a single firm, or barriers to entry preclude competitive responses to high pricing or exclusionary conduct.

An efficient monopoly exists in a particular market where the marginal unit costs of supplying the relevant product in the relevant geographic market are least if there is one supplier, over the longer term, without any artificial barriers to entry or exit in the market. This requires constantly increasing economies of scale and/or scope for the industry as a whole in the market for the entire level of demand in the long term even as technology changes.

These conditions are very difficult to satisfy. Very high, fixed, start-up costs (high economic entry barriers) coupled with a capital-intensive business structure and assets which have little or no alternative use or secondary market value (high economic exit barriers) are often considered as necessary conditions for a natural monopoly. Plant- or firm-level economies of scale or scope are, however, generally not regarded as sufficient for a natural monopoly in a particular market. As well, the more rapid the pace of technological change (and hence the greater the capacity to innovate) and the more physically mobile or generic the principal assets of firms in a market are, the less likely an industry is to meet the economic conditions for a natural monopoly.

High entry barriers can also be created by government regulation, trade policy, government procurement preferences and capital market imperfections.

In other words, the current *Competition Act* and, in my view, probably any form of competition law which aims at preventing consumer welfare losses from private sector restraints to trade on an exceptions basis, is not likely to prevent monopolistic conduct in industries characterized by high concentration and/or high entry barriers. Specifically, the current abuse-of-dominant-position practice is clearly directed at forestalling monopolizing behaviour (behaviour that is directed at reducing competition) in intrinsically competitive markets. It seems almost axiomatic that an efficient monopoly or a firm that becomes more efficient as competition is eliminated could not practice anti-competitive acts in its own market since the elimination of all its competitors would result in production at the lowest possible costs.

Classical economic theory suggests that a firm that profit-maximizes in an industry with very high entry barriers would undersupply the market relative to a situation of competitive supply. The firm would price-discriminate according to consumers' ability to pay in order to appropriate to itself as much consumer welfare as it could.

However, these reservations do not apply to the extension of monopoly or market power from that base market to other intrinsically competitive markets. The *Competition Act* has several reviewable practices specifically tailored to preventing such anti-competitive monopolization.

The second assumption behind competition law cannot readily be met where consumption and production decisions are so fraught with non-price externality effects that prices cannot be used as a benchmark of resource value trade-offs. Equally, markets fail if prices cannot be created or collected. These latter conditions define a pure unpriced public good.

The former externality conditions define situations where market failure may be addressed to a limited extent by what economists call "second-best" solutions through subsidized or mandatory consumption (where prices exceed marginal social value resulting in insufficient consumption) or by the imposition of social cost taxes, mandatory consumption restrictions and/or mandatory product design rules to minimize social costs not reflected in the price (where prices are less than marginal social value).

Safety regulations are an example of this first situation; applied in a market on an evenhanded basis, safety regulation imposes general costs akin to taxes. Such regulation therefore would not distort input prices to a degree that would make competition law impossible to apply because, in practice, adequate evidence could not be developed to support a remedial Order or criminal conviction. Safety regulation and competition law can therefore comfortably co-exist.

Pure public goods may readily be identified based upon the physical inability to exclude non-payers from the benefits of consumption by payers or conversely the physical inability to get users to pay for any socially useful level of production. They include policing and defence and possibly little else. They do not necessarily include goods originally provided by the state to accelerate or increase their production from what the private sector was then offering, but which now can be produced abundantly by private markets or which can now be effectively charged for (that is, potential free riders can now be excluded).

Whether the production or consumption of a non-pure public good is so enmeshed in externalities that prices are meaningless is largely a matter of political judgement.

However, to a significant extent such externalities can be minimized through consumption or product standards or explicit targeted subsidies in order to maximize the ability of individuals to exercise price-based choices. As a general proposition, consumer welfare as a whole would appear to be better served through a system of explicit and targeted subsidies and/or consumption standards for which there is clear political accountability than through a price system involving hidden internal and untargeted subsidies or hidden consumption standards.

D. INSTITUTIONAL/ADMINISTRATIVE DIFFERENCES

1. Introduction

This section examines the institutional and administrative differences that exist between the government bodies that apply the substantive business conduct and industry structure standards of the *Competition Act* and the *National Transportation Act, 1987* and related federal and provincial passenger transport regimes.

As previously noted, there is no statutory or common law requirement that the National Transportation Agency or provincial intercity bus regulators adopt substantive decision-making criteria that *differ* from those of the *Competition Act* in those areas where their jurisdiction either parallels or excludes the operation of that Act. Such bodies apply sufficiently broad discretionary tests that fitting *Competition Act* criteria into them would be lawful even if the regulator had previously applied different criteria. The only requirement in doing so would be adherence to the administrative law rules of fairness or natural justice.

Any difference between the regulators' criteria and the prevailing standards of the *Competition Act* must therefore be regarded not as a legal requirement but as a result of the balancing of competing principles and interests established as legitimate through the individual regulator's institutional structure and/or through the decision making rules imposed upon the regulator by the courts, particularly due process requirements. It is also conceivable that the regulatory agency may employ due process standards which go beyond the minimum level required by the courts. It would do so to satisfy general political expectations respecting direct public involvement in decision making, or to increase the legitimacy of its actions by co-option of affected interests.

2. Regulatory Agencies — A Behavioural Profile

Procedurally, one thing can lead to another. Once having embarked upon a decision-making process that involves identification and ranking of legitimate or affected political or economic interests (a polling process), a regulatory agency will find it very difficult to order its priorities or to structure its decisions in a fashion which does not fall within a "range of reason" established by the positions of the interests involved in its decision-making process. Over time, a regulatory agency may also base its priorities upon the more dominant or persistent economic or political interests appearing before it (these can be producer, consumer or even governmental interests).

Under these circumstances it becomes very difficult for the regulatory agency in practice to implement priorities based upon objectives, such as maximization of competition which, over the long term, are relatively abstract and do not appeal to the direct economic interests that are involved in its decision making.

For example, *future* competitors are, by definition, never directly present in a decision-making process.

This difficulty increases with the vagueness or generality of the statutory decision-making criteria. This considerable vagueness is inevitable, in my view, once a regulatory agency model is adopted where investigation and adjudication are merged in a single institution. The absence of statutory or executive policy guidance and the resulting dominance of jurisprudence and consensus-oriented process also significantly reduce the ability of this form of regulatory agency to adopt dramatically different behavioural or structural standards for an industry to reflect changes in the industry's economic fundamentals.

Because a consensus-oriented, decision-making process involves a degree (often very significant over time) of interest-group brokering, there is often compelling pressure upon the regulatory agency to make decisions that are "fair" to the interests involved. The tendency is bound to increase to the extent that the regulator's enabling legislation does not provide detailed and internally consistent decision-making standards.

Finally, because regulatory decisions, particularly where public hearings are involved, take on some jurisprudential value, and because a closely knit group of producers with strong expectations for consistency of regulatory policy/jurisprudence (the regulated sector) are involved in its decision making, it becomes extremely difficult politically for a regulatory agency to amend significantly its objectives or priorities. Typically the initiative to change objectives or priorities comes from an exogenous "shock to the system," such as dramatically different market conditions affecting the regulated sector, government policy, or (most rarely in Canada) judicial intervention.

Put simply, as the result of a combination of factors, regulatory agency decision making tends over time increasingly to broker competing economic interests through decisions that are politically legitimate (that is, fair). There is generally no internal correction or arresting mechanism in the agency's structure or in its instructions from the legislation to keep the agency from continuing along this slippery slope.

This correction must generally be imposed externally through legislative change (including deregulation). For reasons discussed below, other external options such as judicial intervention, legislative review of appointees or agency performance, executive policy guidance, and staff and decision-maker overhaul (such as is applied in the U.S. revolving-door approach) are not significant in the Canadian context.

It would be useful, therefore, to summarize the convergence of institutional and structural factors which support decisions that gravitate to political fairness and modest change rather than maximum competition and rapid change. These factors are in relation to Canadian regulatory agencies, such as the National Transportation Agency, that combine investigative and adjudicative functions and employ open public hearings in their decision-making process.

(a) Absence of Clear Statutory or Executive Policy Guidance

Tests such as "just and reasonable," "public convenience and necessity," "no undue discrimination," and especially "in the public interest" are no more than large receptacles for the case-by-case exercise of unconstrained administrative discretion. They have no substantive or operational meaning in and of themselves.

Regulatory statutes are seldom revisited by the legislature in a fashion which would cause a reassessment of priorities. Opportunities for executive policy guidance are few and largely unused.

(b) Little or No Policy Accountability of Regulatory Appointments

Appointments are often political rewards and almost never designed to ensure application of a particular economic policy. Appointees are often not selected because of their expertise and where expertise is a criterion, the expertise often is developed through involvement with incumbent producer interests. Reappointments or new appointments occur gradually over time; thus the complexion of regulatory agencies, much like the Supreme Court of Canada, is hard to identify and political balances are slow to change.

Appointees who do not perform effectively are almost never removed for cause before the appointments period ends. Rather they are not re-appointed. Few proposed appointees are even interviewed publicly by the legislature

before the appointment is made, to identify either expertise or priorities. The legislature seldom reviews the substance of the agencies' decisions other than during brief examinations of financial requirements.

Notwithstanding the virtually unfettered review and override powers of Cabinet, and thus the opportunity for Cabinet to articulate in advance the circumstances under which overrides may occur, extremely few Canadian regulatory agency decisions are changed by the executive branch of government.

(c) Longevity of Staff Advisors

In contrast to the fairly even and rapid turnover of agency appointees, Canadian regulatory agency staff tend to be employees of long duration with senior staff often recruited from the regulated sector itself or other branches of government that regard the particular regulated sector as a client from an economic development perspective. This difference tends to increase the influence of staff in the decision making and provides an important opportunity for the regulated sector to exercise decision-making influence outside formal or public elements of the process.

As well, the agency staff, rather than agency appointees, tend to be the source of industry expertise.

(d) Absence of Judicial Intervention

Typically, judicial review of agency decisions is restricted to errors of law or jurisdiction, which can include insufficient procedural fairness, misinterpretation of statutory decision-making criteria, bias and acting for an improper purpose.

However, given the inherent vagueness of regulatory decision-making criteria and statements of legislative policy in Canadian regulatory statutes and the absence of constitutional economic rights, there are virtually no opportunities for judicial intervention into the merits of a decision.

As well, Canadian due process law respecting regulatory agencies is relatively new and costly and time-consuming to activate. It has to date not provided as effective a tool as American due process law in ensuring a balance

between competing economic interests (for example, producers versus consumers) in access to regulatory information or access to decision-making activities (including off-the-record influence).

(e) Role of Public Hearings

Public hearings can be a device for levelling the impact of resource disparities among competing economic interests. They also can be a double-edged sword. Public hearings open up at least some part of the overall decision-making process to a structured competition between the articulated merits of competing economic interests. On the other hand, public hearings are costly and, if there is no tradition of open access, public hearings can reinforce the impact of resource disparities between producers and consumers and favour advancement of the views of the traditional dominant player, usually the regulated sector as a whole or major incumbent firms.

The value of public hearings as a levelling device also decreases where their actual use or relative importance in the decision-making process cannot be predicted with accuracy. For example, regulatory agencies often have considerable discretion over whether or not to conduct a hearing in order to decide, and, if a hearing is conducted, considerable discretion over its role remains. In practice, if an agency's decision does not expressly state that the record of a hearing was ignored in making the decision, judicial intervention into the agency's reasons is highly improbable. The agency is thus free, in relation to the risk of judicial intervention, to ignore what might reasonably be supposed to be a key, if not determining, decision-making event.

3. Transportation Regulatory Agencies

Several considerations affect an assessment of the comparative capacities of direct regulatory intervention and competition law to protect consumer interests that relate to specific design features of economic regulators for Canadian passenger transportation services.

First, all such agencies have their genesis in a legislative presumption that the sector requiring regulation was subject to chronic and significant market failure. There are no deregulation provisions in the enabling legislation to permit the regulator to bail out of counter-productive intervention in the event that the circumstances causing market failure disappear. This neatly avoids any need by the regulator to review the underpinnings of its continued

existence or, in the extreme, to consider self-destruction. Rather, in practice, it provides an opportunity for the regulator to fortify its continued existence by consciously promoting the continued validity of the presumptions that led to its existence in the first place.

The market failure prompting regulatory legislation appears to have differed among modes. For rail, it was a fear of natural monopoly power. For air and trucking, it was a fear that unconstrained competition was inherently destructive and would result in inadequate levels of service, rapid swings between over- and under-capacity and few incentives to invest. It is now recognized that, to a large extent, market failure rationales played a secondary role to the sheer political power of producer interests and were largely developed as after-the-fact ways of legitimizing interests wishing to maintain a regulated industry structure.

The concern over inadequate service levels also, before and after the fact of agency creation, provided a strong opportunity to generate a regulated sector price structure that was riddled with hidden but politically attractive internal subsidies. These subsidies reduced or eliminated the need for explicit government subsidies covering the start-up phase of the sector, or individual firms, avoided the uncertainty and political risks of their possible removal, and permitted the regulator to appear fair to competing consumer interests.

Finally, because transportation technologies developed at different times and the political balance of the federal and provincial levels of government has shifted over time, transportation regulation legislation has been historically confined to single modes (that is, technologies). Jurisdiction has been split between the federal and provincial levels based upon political rather than economic considerations. Rail and air services are federally regulated while road-based services are effectively provincially regulated. Public road passenger jurisdiction is also split at the provincial level between the province itself (buses) and its municipalities (taxis).

This historical accident has created an inherent bias within individual transportation regulation bodies to regard each mode as a discrete industry having a discrete market rather than functioning as a substitute for other forms of public or private transportation.

However, mode-specific analysis tends to avoid the real questions to be answered in identifying continuing market failure such as: given all available substitutes in providing transportation services within a particular geographic area to a discrete set of consumers (for example, passenger service consumers), is it possible for any one supplier or group of suppliers (without an agreement not to compete) to earn monopoly profits or to stay in business while providing an unacceptable level of service?

Mode-specific analysis also tends to discourage an examination of the impact of enforcing traditional hidden subsidies, through regulated prices in a particular mode, on the long-term survival prospects of that mode where technological change offers new effective substitutes.

4. Competition Law Contrasted

(a) Substance

There are major differences in the substantive conduct standards of competition law and the direct regulation model and in the manner in which these standards are applied. These differences are critical to an assessment of the relative merits of each approach for the future protection of intercity passenger service consumer interests. They may be summarized as follows:

- (i) Statutory industry performance standards are more explicit under competition law.
- (ii) Competition law remedies are designed to be transitory. Firms or industries subject to competition law intervention and restraint have this restraint lifted by the courts or the Competition Tribunal once the statutory prerequisites for restraint disappear or circumstances change.
- (iii) Markets, and hence the presence of market failure, are defined in a consistent manner for all aspects of the economy free of institutional presumptions or any market analysis of the presence of functional substitutes and the prospective impact of competitive entry regardless of the technology employed to produce the final product. This, in my opinion, makes competition law, as a general law of general application, a more effective instrument for assessing the effects on a particular sector or industry of evolving market or technological conditions. Competition law also encourages an analysis of the consumer impact of firm or

industry behaviour which emphasizes the likely response of competitors or investors to that behaviour over the current performance or structure of an industry;

- (iv) Competition law precludes intervention to generate hidden consumer subsidies through price structures and therefore places the subsidy responsibility directly on the appropriate legislature. The objectives of economic efficiency, industry development and wealth distribution are thus not commingled in a murky jurisprudential soup that militates against public accountability over the adoption of often non-complementary policies;
- (v) Finally, in my view, the overall set of actionable behaviour under Canadian competition law is largely congruent with the behaviour described by established microeconomic theory as most likely to reduce consumer welfare through higher prices, reduce choice or reduce innovation. In contrast, the open-ended nature of transportation industry regulatory intervention standards creates an opportunity to make actionable behaviour which could enhance consumer welfare. A possible exception to this observation is, as noted previously, a situation of natural monopoly.

(b) Process/Administration

There are also very basic differences between competition law and transportation regulation administration. Competition law is enforced equally against all elements of the economy with the investigative or fact-finding process occurring in private without there being any enforceable involvement by affected economic interests. Competition laws are adjudicated in a strict judicial environment even where the Competition Tribunal is the adjudicator. The adjudicator is exposed only to a series of "one-off" issues relating to specific behaviour of a specific firm.

On the other hand, transportation regulation is enforced only with respect to specific industry components with fact finding involving some enforceable input rights for affected interests. Issues are adjudicated in a less judicial and more political environment. The ultimate decision maker is exposed to what amounts to a continuum of issues relating to the business policies of that particular industry component.

These differences, in my view, have resulted in compelling incentives for transportation regulators to go beyond the level of intervention necessary to maximize consumer welfare. They should become a "second layer of

management” which attempts to strike a delicate balance, within a non-managerial environment, between competing suppliers, consumers and industry development goals. They should use a statutory framework that encourages decisions to be made on the grounds of fairness. The decision-making structure of transportation regulation, as with other industry-specific regulatory regimes which constrain competitive industries, contains a variety of elements which collectively create a bias in favour of protecting incumbent supplier interests. This ultimately leads to a less than fully efficient development path for the industry as a whole.

(c) Conclusion

The substantive and administrative structure of competition law appears to provide reasonable assurance that consumer welfare will be maximized in the supply of transportation services, as with other services, with the exception of those transportation services that are demonstrably natural monopolies. There is a built-in tendency of transportation regulation to import and balance, largely in a hidden unaccountable fashion, policy considerations which are different from, or in conflict with, consumer welfare maximization. Competition law in my view is a superior instrument to transportation regulation to maximize consumer welfare, and therefore to protect consumer interests.

An exception to this general proposition could be made where regulatory protection was determined to be an appropriate transitory policy during the infancy period of a new industry (to be “sunsetting” once a sustainable and competitive industry structure was achieved). However, on balance, regulatory protection would appear to be an excessively risky instrument to distribute wealth among consumers. There are too many serious industry inefficiencies caused by the prices and entry barriers necessary to sustain cross subsidies.

V. TRANSPORTATION INDUSTRY DEVELOPMENTS: APPROPRIATE TRADE PRACTICES FRAMEWORK

A. INTRODUCTION

This part examines the suitability of competition law regulation of trade practices, and the likely approach of the Bureau of Competition Policy towards complaints, in relation to the following industry developments:

- (1) increased cross-modal ownership (for example, bus and air feeder services);
- (2) infrastructure access arrangements and pricing;
- (3) airport defederalization;
- (4) road privatization;
- (5) intercity bus service deregulation;
- (6) air carrier entry and price deregulation in Northern Canada;
- (7) increased competition from non-Canadian suppliers (for example, fifth freedom or cabotage rights provided to U.S. carriers; increased airport access under bilateral arrangements); and
- (8) passenger ferry privatization.

B. INDUSTRY DEVELOPMENTS

1. Cross-Modal Ownership

A competition law analysis of industry structure and concentration of ownership changes is conducted in relation to the relevant product and geographic market rather than changes in the relative financial or bargaining power among firms engaged in a particular transportation mode. Such an analysis would consider the nature of economic and regulatory entry barriers in each mode and whether it was likely that increased cross-modal ownership would, on a case-by-case basis, foreclose single-mode competitors from important sources of customers and interconnecting traffic.

There would only be a potential competition law concern if the respective modes supplied the same markets or if one mode was an essential input to another market. No concern would arise if there were no apparent economic or regulatory barriers to entry into the relevant market. Economic barriers to entry would be examined in relation to the extent of fixed and start-up costs, and the capacity of customers to switch suppliers.

Given the low economic entry costs in most transportation services, the mobility of transportation assets (excluding roadbed which is generally separately owned), and the maturity of development of the passenger transportation sector, there is little reason to conclude that the basic set of

trade practices controls of competition law would not satisfactorily prevent industry behaviour that did not result in maximization of consumer welfare.

Without regulatory entry barriers, there is little chance of an actionable competition law issue arising. This further supports the relative desirability of the less interventionist administrative structure of competition law.

2. Infrastructure Access Arrangements and Pricing

This next section examines the usefulness of the *Competition Act* in constraining anti-competitive behaviour with respect to controlling access to essential “bottleneck” inputs to the production of passenger transportation services. These inputs include access to airport terminal facilities, railway, roadbeds and highways.

Before proceeding with this discussion, it should again be emphasized that the *Competition Act* itself does not permit the application of social, environmental or industrial development policy goals on economic decision making. It takes the prevailing price system or conditions of supply as given. To the extent that these prices or conditions of supply reflect such policies, the *Competition Act* becomes an increasingly less effective instrument in encouraging economic efficiency and consumer welfare.

This inverse relationship should, of course, not be surprising since it essentially reflects the overall balancing being made by society through its laws. Economic efficiency as measured purely by private transactions and their prices is balanced against overall social welfare, the measurements of which become permeated with subjective and political judgements because of the presence of important perceived, but not easily measured, consumption and production externalities.

For example, assume that roadway usage charges for large, energy inefficient passenger cars were set by a supplier well above the long-run marginal costs imposed on the roadway by such cars in order to discourage their consumption and thereby to reduce pollution. It would be far more difficult for the Bureau of Competition Policy to examine a complaint from a manufacturer or rental car supplier that the pricing practice had anti-competitive effects (for example, in relation to competition with public transit including intercity buses, or smaller car suppliers) than if the roadway usage prices had been set using the principles of marginal cost pricing.

The following discussion will therefore assume away pressure favouring substantial “social policy” pricing in the commercial supply of transportation infrastructure access.

Two situations are considered: (1) the supplier of infrastructure access is also a supplier of passenger transportation services which require the infrastructure or which compete with services that require that infrastructure (a vertically integrated supplier); and (2) the supplier of the infrastructure is *not* vertically integrated with a supplier of passenger transportation services.

In both instances, assume that the infrastructure supplier is the sole supplier in a geographic market but that there is no regulatory licensing barrier to prevent or control new infrastructure suppliers in that market.

(a) Vertically Integrated Supplier

This situation could arise, for example, if

- (i) roadway access was supplied by a commercial Crown corporation or an investor-owned firm which also controlled or had a significant investment in suppliers of public transit, intercity buses or taxi-type passenger transportation services;
- (ii) rail access was supplied by a commercial Crown corporation or an investor-owned firm which also controlled or had a significant investment in a passenger rail service or intercity bus service supplier; or
- (iii) airport access was supplied where the airport owner was a commercial Crown corporation or an investor-owned firm which also controlled or had a significant investment in an airline.

The *Competition Act* contains several quite clear provisions directed at preventing vertically integrated suppliers with strong market power in one market from extending that power through exclusionary or discriminatory pricing practices into a second, more competitive market made up of unintegrated suppliers or customers.

In situations where the vertically integrated infrastructure supplier was attempting to control access through discriminatory pricing against third-party customers, discriminatory access conditions, or outright refusals to supply access, the following provisions would apply:

- (i) *refusal to deal*: Assuming infrastructure access is supplied on a commercial basis, then the Competition Tribunal could order the infrastructure owner to supply access to parties on usual trade terms if there is a refusal to supply but access is in ample supply, and the third-party passenger service supplier is substantially affected or precluded from carrying on business due to an inability to obtain adequate supplies. This inability must result from insufficient competition, and the third party must be ready, willing and able to meet the infrastructure supplier's usual trade terms;
- (ii) *abuse of dominance*: The illustrative anti-competitive acts set out in the *Competition Act* include "squeezing, by a vertically integrated supplier, of the margin available to an unintegrated customer who competes with the supplier for the purpose of impeding or preventing the customer's entry into, or expansion in, a market" (this is aimed at dominant supplier price discrimination), and "pre-emption of scarce facilities or resources required by a competitor for the operation of a business, with the object of withholding the facilities or resources from a market" (this is aimed at discriminatory supply conditions by either a vertically integrated or unintegrated supplier either to inhibit competition or to obtain monopolistic profits).

As previously noted, dominant position and the effect of a substantial lessening of competition from the practice of anti-competitive acts must also be proven using a civil law standard of proof.

There is, as has been discussed, a considerable jurisprudence under the *Competition Act* dealing with exclusionary practices and pricing by vertically integrated suppliers. In some cases, such as the supply of replacement Chrysler auto parts or aspartame sweetener, the supplier, subject to a Competition Tribunal review application was, in effect, a monopoly.

Clear applicability and orientation of the *Competition Act* are not in doubt. The question is whether the Act and its administrative apparatus provide a sufficiently practical and timely way of remedying this form of anti-competitive behaviour.

It would be inappropriate to focus only upon the average turnaround time of litigated cases before the Competition Tribunal that ultimately result in a Tribunal remedial Order. The bulk of complaints arising under the reviewable trade practices provisions of the Act are settled informally, after Bureau

intervention outside the formal inquiry process, to the apparent satisfaction of complainant and supplier. This suggests that timely remedies on an informal basis are the norm. As well, suppliers are increasingly relying upon opinions of the Bureau, provided under its Program of Compliance to structure their businesses.

One interesting option may therefore be, at the time at which infrastructure access is to be put on a commercial basis, to require, as a condition of business transfer to the new supplier, that the new supplier obtain and adhere to a Program of Compliance opinion on how to avoid *Competition Act* violations.

With respect to the practicality of the remedy that may be provided under the *Competition Act*, the refusal to deal and abuse of dominance provisions should be distinguished. The refusal to deal remedy is resumption of supply on normal order terms — it does not allow for an offset for advantages obtained from past anti-competitive behaviour. However, abuse of dominance provides a more sweeping power to the Tribunal if a cease and desist order is not likely to restore competition: to order whatever is reasonable and necessary to overcome the effects of the anti-competitive practice in the market, *including* divestiture of the firm against which the Order is directed.

(b) Non-Vertically Integrated Supplier

As previously discussed, the abuse of dominance position provisions of the *Competition Act* are, in my view, not well suited to preventing market failure in markets where a single supplier is most efficient or where there are extremely high economic or indirect regulatory entry barriers. Although the abuse of dominance provisions do include a divestiture power, this power is, from a practical standpoint and in relation to the overall remedial structure of the Act, a power which would be seriously considered only in exceptional circumstances. Moreover, divestiture of a supplier which is, in effect, a geographical natural monopoly may not be a complete answer if the only result is the creation of a number of smaller geographic natural monopolies.

As well, the overall orientation of the *Competition Act* is towards the remedy of transitory problems in inherently competitive markets.

Consequently, the situation of the unintegrated monopoly supplier should be examined both in terms of its home market and in terms of its impact as a supplier of inputs to competitive retail level passenger transportation services.

The same analysis as applied to vertically integrated infrastructure suppliers would apply to the unintegrated supplier with respect to price or supply discrimination among customers in relation to commercially available products. The motive and consequences of anti-competitive behaviour would, however, become substantially more difficult, although certainly not impossible to establish if the infrastructure supplier did not have a direct economic interest in a transportation service competitor. However, the case could still be based upon the comparative impacts of discriminatory prices that do not reflect marginal costs or discriminatory supply restrictions on competitive firms.

Both the evidentiary problems and conceptual orientation of the abuse of dominance provisions become significantly greater when inter-competitor discrimination is not an issue, but instead, the complaint arises from perceived inefficiency, excessive prices and profits, or absence of innovation on the part of the unintegrated sole supplier. In short, this is the traditional problem of monopolistic profits, goldplating, and dulled performance that has supported both government ownership and public utility regulation.

Previously noted was the illustrative anti-competitive act of “pre-emption of scarce resources” which conceivably could be extended to monopoly under-supply of a market (which classical economics theory suggests will occur with a profit-maximizing monopoly). However, in my view, this is stretching the scheme of abuse of dominance as presently written.

As well, it appears uncertain whether there is an effective remedy available to the Competition Tribunal in these circumstances. The Tribunal has no authority to control prices or profits or to require the provision of yet-to-be supplied products. As noted, the divestiture power is not likely to be any help in these circumstances.

Thus, in my opinion, either amendments to the *Competition Act* or a separate regulatory regime would be necessary to control the monopoly power of an infrastructure supplier that was not manifested in discriminatory pricing or supply conditions among its customers.

On the other hand, it does not necessarily follow that the only real options are government ownership (the effectiveness of which is not examined in this report) or direct (public utility-type) regulation of services, profits and products.

For example, the infrastructure supplier's pricing behaviour could be made subject to the *Competition Act*, but its overall corporate performance could be subject to non-regulatory incentives such as a periodically renewed franchise tied not to the continued operation of the business but to management of the business. A remedial trusteeship of specified duration could be imposed by government to supplant incumbent management if profits, costs or the level of overall innovation appeared unreasonable upon a periodic corporate performance review. Another option, which arises at privatization, is the retention of a "golden share" in the hands of government which, coupled with a shareholder's agreement, requires government or third-party review of overall corporate performance, but not specific pricing and supply decisions.

Such options would, in my view, not cast a cloud over the applicability of the *Competition Act* to pricing and supply decisions. To assess their effectiveness, they should be compared to direct regulation of access prices and conditions by an independent agency.

The principal benefit of direct intervention is that government control of industry behaviour would be regarded as fairer and less-hidden since decision making would occur in a more public and consensus-oriented forum. Access prices and conditions certainly, or stability, if they are considered to be important outcomes, do not require direct regulation to be realized since price transparency and minimum-price duration requirements could readily be made through franchise renewal or divestiture conditions.

On the other hand, assuming that monopoly profits can be precluded through franchise renewal or divestiture conditions, the regulatory agency option, as it has developed in practice in Canada, runs the risk of establishing prices that are no more conducive of efficient economic behaviour than is current rationing of capacity on non-price basis. A regulatory agency is also a costly administrative apparatus that adds considerably to the complexity of supplier decision-making.

To some degree, this problem of natural monopoly performance could also be mitigated if, in the course of infrastructure privatization, the infrastructure was broken up into a substantial number of distinct geographical monopoly suppliers. To operate, these suppliers would have to enter into traffic exchange or joint operating arrangements with each other in order to be efficient.

Two conditions would prevail: (1) these suppliers would have strong financial incentives to maximize traffic and obtain access to other infrastructure suppliers at minimum cost, or (2) there would be consumer transparency among infrastructure suppliers.

Condition (1) could arise, for example, in a privatized road system where users were billed by one "home-base" system for all travel, but could shop around for the most economic home-base supplier. Condition (2) could arise where all road services were billed on a usage basis, and the prices of all road service suppliers were separately identified on each bill.

Such a market structure could impede monopoly practices through the balancing of bilateral local monopoly bargaining power. It might also produce sufficient price comparisons of firms with similar cost structures and comparable (but not identical) terms of supply to permit an analysis of whether any one firm might be subject to a remedy under abuse of dominant position.

Here, however, an effective anti-conspiracy control (again possibly taking the form of a binding Bureau Compliance Opinion) would be essential to ensuring the establishment of useful "shadow market" information, and to prevent industry-wide coordination practices.

3. Airport Defederalization

The discussion in (b) applies to the supply of airport facility access. Airport access is presumably supplied on a quasi-commercial basis. The constraint to current application of the *Competition Act* is ownership: airport access is provided by a federal department, the supply activities of which remain exempt from the Act by operation of the common law on Crown Agent coverage.

If airports were sold to private investors or a Crown corporation of either the federal or provincial governments, there is no doubt that the *Competition Act* would then apply. If airports were sold to municipal governments, the Act would also apply since such governments are statutory creatures. However, municipalities have the capacity to become Crown Agents if expressly made Crown Agents by legislation.

Notwithstanding that airports are presently subject to exclusive federal regulatory jurisdiction, it still appears possible for the provincial Crown to

make airports Crown Agents thus exempting them to the extent possible from the *Competition Act* until that Act is amended to cover all Crown Agents.

4. Road Privatization

In addition to vertical integration and natural monopoly issues discussed in subsection 2 above, road privatization also raises the issue of the capacity to price, bill and collect on bills — an issue that in effect defines whether roads should continue to be regarded as potentially private goods, as opposed to the current view that roads are a pure public good.

Without an actual pricing system for roads, even with privatization to a firm subject to the *Competition Act*, the capacity of this legislation to prevent behaviour that reduces consumer welfare would be low. However, it could apply to contractual access terms which could constitute abuse of dominance, for example, measures directed at frustrating competitive entry.

5. Intercity Bus Service Deregulation

The only serious barrier to intercity bus service deregulation is the presence of a significant degree of geographic cross-subsidization which is considered to be necessary to maintain the current geographic coverage of bus service. This requires an assessment of barriers to entry and probably pricing of new entrants services in “thin markets.”

This industry segment is not a natural monopoly. Buses, like freight transport trucks, are low-cost, highly mobile and self-sufficient factories. On this basis it would appear that the *Competition Act* provides a sufficient trade practices regime for this industry.

6. Air Carrier Entry and Price Discrimination in Northern Canada

Again the effectiveness of competition law as a consumer protection instrument should be examined in light of the structure of the air services sector under this regime, particularly entry barriers, and the availability of modal substitutes.

Clearly the availability of modal substitutes to passenger air services is less in Northern Canada than in Southern Canada. However, the capital cost associated within Northern route entry appears to be quite modest in

relation to thick interurban route entry in Southern Canada since many Northern markets can be well served with small propeller-driven passenger aircraft. As well, passenger service in Northern Canadian routes would appear to be virtually a joint product with the supply of cargo services, since a relatively high proportion of goods consumed in Northern communities (and particularly high-value goods) is shipped in via air. Thus there would appear to be quite low barriers to entry into supplying air passenger services to the region as a whole and to particular origin-destination pairs in that region.

On the other hand, given the overall importance of air transportation to such communities, and the vulnerability that comes with remoteness, security of supply becomes a more important element of the overall consumer interest than with respect to other passenger service modes. Even a service disruption of as short a period as a week can be of major concern to Northern communities. This is the basis for the entry and exit restrictions established in the *National Transportation Act, 1987*.

Under Canadian competition law, competitiveness in terms of rate of new entry has traditionally been measured in longer time frames than are likely to be tolerated by Northern Canadian consumers of passenger or cargo air services. However, if a solution were to continue to be provided with respect to security of supply, it would appear that, given the overall inherent competitiveness of this air transport industry segment, competition law could be satisfactorily substituted for the other direct regulatory constraints of the *National Transportation Act, 1987*. A mandatory exit waiting period as a condition of an operating licence granted under a "fit, willing and able" test could be sufficient. A supplementary measure, if warranted, might include a security deposit or bond to permit community chartering of replacement service in the event of a breach of this waiting period condition.

Finally, as a general proposition based in part on airline deregulation in the U.S., particularly Alaska, a program of direct thin-route subsidies may be a low-cost and easily administered option for maintenance of supply security.

7. Increased Competition from Non-Canadian Suppliers

This issue may be considered in relation to both air and bus passenger services. However, the geographic coverage of U.S. or overseas air service

suppliers is clearly greater. And the most immediate prospect of competitive entry is from U.S. carriers having a significant existing transborder business.

The principal issue for consumer protection is whether foreign-based suppliers would have an incentive and an ability to enter the Canadian market at low prices, force Canadian incumbents out of the market, and then raise prices to levels that would otherwise not be possible.

More succinctly, the issue is the likelihood of successful foreign supplier predatory pricing, or effective service “dumping,” to use the international law equivalent.

First, it should be noted that the predatory pricing, price discrimination and abuse of dominance provisions of the *Competition Act* apply to all prices charged in Canada by persons engaged in a business in Canada. Thus, it would be incorrect to assume that, because the *Special Import Measures Act*, which contains Canada’s anti-dumping regime, does not apply to the importation of services, there is no Canadian law applicable to predatory pricing of service imports. The *Competition Act* applies to the supply of such imports in Canada.

Any uncertainty as to whether a foreign-based passenger services supplier was engaged in a business in Canada could adequately be overcome by:

- (i) requiring an identifiable presence in Canada as a condition of supplying Canadian origins or destinations;
- (ii) requiring that business presence to provide adequate security by way of a performance bond or other security deposit; and
- (iii) making compliance with Canadian pricing laws a condition of entry and continued Canadian business rights.

It is in this context, that consumer protection measures beyond the *Competition Act* should be examined.

Prerequisites for effective predatory pricing are:

- (i) capacity in other markets to raise prices above competitive levels to fund entry through predatory prices; and
- (ii) a capacity to withstand further entry in the market where predatory pricing has been practiced to reduce or eliminate incumbent competition.

This requires an analysis of the level of competition and barriers to entry in both the established and the new market. With respect to the U.S. air transport sector, there is a very high degree of excess plant capacity (unused operating aircraft). This strongly suggests that there is little or no incentive or capacity on the part of U.S. air carriers to use revenues in the U.S. market to fund sustained below-incremental cost entry into Canada since both the "home base" and the Canadian markets are intrinsically highly competitive and have low economic entry barriers. Equally, the low entry and exit barriers in air transport would strongly suggest that any post-predation attempt to raise prices above competitive levels would be quickly met by price competition.

The only possible basis for effective U.S. based, air service predation into the Canadian market would be an agreement among all U.S. suppliers authorized to enter the Canadian market not to compete among themselves domestically and in Canada in a fashion that would permit post-predation joint monopolization of the Canadian market. The agreement not to compete in the U.S. to facilitate this export cartel strategy would of course be subject to U.S. anti-trust laws which are much more severe than the *Competition Act*. The agreement would also be subject to the conspiracy prohibition of the *Competition Act*. These exposures would have to be weighed against the prospective permanent benefits of a U.S.-supplier, market-sharing arrangement governing the Canadian market.

This joint monopolization scenario appears very remote. Not only is an effective joint monopoly unlikely, but, in my view, U.S.-based suppliers are unlikely to risk U.S. anti-trust sanctions and the prospect of being shut out of the Canadian market under a treaty arrangement as part of a Canadian government response.

Accordingly, the remote likelihood of successful predation as part of further U.S. air service entry supports reliance on competition law as the

appropriate trade practices regulation instrument with respect to U.S. air services competition.

Basically the same analysis applies to support reliance on competition law with respect to U.S. bus service entry.

This analysis, however, may not apply with equal force to further entry by non-U.S. based air services since these services are by-and-large national monopolies or obligations, and entry into these foreign national markets is often very strictly controlled. Thus, there may be a domestic market structure that might support successful predation. However, for such services the question must be asked whether any such national carrier would have a real business incentive to practice predatory pricing in the Canadian market, especially since such pricing behaviour would be very easy to detect quickly. In all likelihood, there would not be a greater incentive or capacity to predate successfully by such suppliers.

The *Competition Act*, therefore, would probably provide the appropriate regulatory framework for non-U.S. air service entry as well.

8. Passenger Ferry Privatization

The same considerations respecting natural monopoly power as were discussed in subsection 2 in relation to infrastructure access arise also in relation to marine ferry privatization.

The *Competition Act* as presently structured would not readily provide an adequate restraint to the exercise of market power unless competition were introduced through privatization. This could be through the sale of ferries serving a route to more than one supplier, coupled with a fairly strict anti-conspiracy control (possibly implemented by the Bureau of Competition Policy under its existing program of negotiating informal compliance undertakings).

In the event that all ferries serving a market were sold to one firm, there may be a need for some form of public-utility type of price and service regulation. This would be particularly true if the sale included the docking facilities or a franchise in the market.

VI. CONCLUDING REMARKS

Part IV examined the substantive, structural and administrative differences between the *Competition Act* and the general characteristics of transportation regulation legislation and regulatory agencies.

The analysis has largely focussed upon the general institutional characteristics of economic regulation as applied in Canada through independent regulatory agencies and the legal and institutional characteristics of Canadian anti-trust or competition law enforcement.

It is not possible, in my view, to conduct an empirical comparative analysis of the economic results of the decisions of the Bureau of Competition Policy and the National Transportation Agency (the Agency). There is a dearth of jurisprudence under the *Competition Act*. But this is even more so with respect to the *National Transportation Act, 1987* since there have been only very few litigated cases producing reasons with respect to pricing, entry, exit and merger of passenger transportation firms — or all transportation firms under the Agency's jurisdiction for that matter. There are no "smoking guns" of a good or bad application of economic principles under either jurisdiction to date.

In any event, for policy-making purposes, a more appropriate focus, is upon the long-term consequences of applying certain combinations of legal standards and institutional delivery mechanisms.

Transportation regulation regimes, due to a combination of factors, are less likely to encourage industry behaviour which is consistent with maximizing consumer welfare over the longer term when compared to competition law. Exemptions to this finding can be made, however:

- (1) where an industry or a particular component of it is in its early stages of development and requires some protection from competition to overcome capital market inadequacies or to maximize the locating of a particular industry in Canada;
- (2) where there are no transactions or transaction prices in the supply of a product; and

(3) where in the market under consideration there is a single supplier, high entry barriers, and the concern is excessive prices or profits.

Competition law applies a clearer and more rigorous set of rules in the prevention of trade practices which reduce consumer welfare. These principles are consistent with prevailing economic analysis of practices that are most likely to reduce consumer welfare.

But, as noted, competition law in general, and the *Competition Act* in particular, is not well suited to providing an effective remedy to the market failure known as “natural monopoly” or where pure public goods are supplied.

The *Competition Act* does not presently cover commercial activities of the Crown (as opposed to Crown corporations).

The *Competition Act* also remains in an evolutionary state due to uncertainty over the constitutional underpinnings of the key conspiracy and merger provisions and continuing dispute over the appropriateness and economic rationale for criminal prohibitions against price discrimination.

The *Competition Act* and its administrators presently suffer a “legitimacy” disadvantage in a democratic society where enduring regulatory legitimacy depends upon some degree of openness of administration and even public involvement in decision making. Presently all inquiries under the Act must take place in private, and public involvement in adjudication is strictly limited by judicial and Competition Tribunal criteria in granting standing to intervene. This impediment to establishing ongoing legitimacy is made all the more apparent when competition law administration is compared to the regulatory environment which involves more open fact-finding and adjudicative processes.

Nevertheless, the Bureau of Competition Policy has slowly but consistently tested the limits on its fact-finding activities and has become more forthcoming in the presentation of its analysis of informally settled matters. It would be possible to relax the legislated privacy requirement without undermining the objective of maintaining confidentiality. For example, the Bureau could be required to present more details on the results of its inquiries in its Parliamentary Report without naming the subjects of each inquiry.

In addition, there is increasing business reliance on, and respect for, the provision of detailed compliance opinions as part of the Bureau's Program of Compliance. Such proactive measures will become increasingly available and useful as the shift from a criminal law to a civil law environment for enforcing Canadian competition law is implemented in practice.

CHARTS

The following charts provide a snapshot of the current jurisdictional coverage of the *Competition Act* and the *National Transportation Act* and underlying provincial bus regulation discussed in Part IV and the relative merits of competition law and direct economic regulatory law in addressing the future developments review in Part V.

Chart 1:
COMPARATIVE JURISDICTIONS: COMPETITION ACT VERSUS PASSENGER TRANSPORTATION AND PROVINCE, PART IV

	<i>Competition Act</i>	NTA, Provincial Bus Regulation
Mergers	x	x
Predatory pricing price discrimination	x	x (busing, rail passenger only)
Agreements to lessen competition	x	
Vertical restraints (refusal to supply tied selling, market restriction, exclusive dealing)	x	
Industry structure level of competition (entry, exit, terms of service)		x
Abuse of dominance (monopolizing conduct)	x	
Excessive prices (monopoly profits)		x

Chart 2
ASSESSMENT, PART V

	Competition Act	Direct Regulation	Possible Other Measures
1. Increased Cross-Modal Ownership	current provisions adequate	parallel jurisdiction not necessary	
2. Infrastructure Access and Pricing	for (a) to (c) Act requires amendment to cover all Crown Agents	generally a second-best solution	linking Bureau compliance opinion with dissolution or privatization
(a) Customer Discrimination	abuse of dominance and non-price vertical restraint provisions adequate	not desirable — high risk of introducing non-economic "fairness" criteria	
(b) Supply Restriction	no means of requiring new products or increased overall supply	no means of requiring new products or increased overall supply	<ul style="list-style-type: none"> • periodic franchise review • golden share • break up territorial monopolies into smaller units • increase infrastructure supplier interdependence subject to strict anti-collusion control
(c) Excessive Prices/Profits	no effective remedy if market supports only one firm	public utility regulation of prices and profits feasible but carries significant risks of dulling innovation and technological change	see 2(b)
3. Airport Defederalization*	amendments to cover all Crown Agents desirable	not necessary (see 2)	creation of a private gate-and-landing-time market subject to <i>Competition Act</i> compliance opinion and undertakings
4. Road Privatization	See 2	See 2	See 2
5. Intercity Bus Deregulation	provisions adequate — industry has low entry barriers	not necessary if very large geographic cross-subsidies are not required	

Chart 2 (cont'd)
ASSESSMENT, PART V

	Competition Act	Direct Regulation	Possible Other Measures
6. Air Carrier Deregulation (Northern Canada)	provisions adequate		<ul style="list-style-type: none"> retention of exit notice period direct thin-route subsidy
7. Non-Canadian Competition	provisions adequate	retain only if domestic industry protection is more important than efficiency or consumer welfare	<ul style="list-style-type: none"> remove 25% foreign ownership limit liberalized entry under bilaterals
8. Passenger Ferry Privatization (a) monopoly profits (b) price discrimination (c) service quality	not adequate not adequate not adequate	second-best solution compared to periodic franchise review review and price limit	periodic franchise review with price cap condition

* This section relates only to transfer of airport ownership and operating responsibility from the federal government to other governments or to the private sector. Controls for monopoly power abuse are examined in 2 and are relevant regardless of the ownership or operating structure for airports.

ENDNOTES

1. *R. v. Carnation Co.* (1969), 4 D.L.R. (3d) 133
 - no practice where prices are a temporary expedient to meet competition.

R. v. Consumers Glass Co. (1981), 124 D.L.R. (3d) (Ont. H.C.J.)

 - prices are not unreasonably low if accused is minimizing losses where there is excess industry capacity.

R. v. Hoffmann-La Roche Ltd. (No. 1 and 2) (1981), 125 D.L.R. (3d) 607 (Ont. H.C.J.)

 - what is unreasonably low is a question of fact depending on all the circumstances.
2. In September 1991, the Bureau published revised draft guidelines. These guidelines do not differ substantially from the April 1990 draft examined in this report, with the exception that the more recent draft does not so explicitly propose a two-stage screening process.
3. This definition is taken from 1990 Draft Merger Enforcement Guidelines, which have been superseded by the Merger Enforcement Guidelines, Information Bulletin No. 5, March 1991 which adopt essentially the same features.
4. Consumer and Corporate Affairs Canada, *Information Document on the Proposed Acquisition of Wardair Inc. by PWA Corporation*, CCAC No. 189 10234E, (Ottawa, April 1989), p. 3.
5. *Ibid.*, p. 4.
6. *Ibid.*, pp. 4–5.
7. *Ibid.*, p. 6.
8. *Ibid.*, p. 6.
9. *Reasons for Consent Order Dated July 7, 1989*, Competition Tribunal, CT-88-1, *DIR v Air Canada*, et al., pp. 69–70.

CONTROLLING MARKET POWER IN WEAKLY CONTESTABLE CANADIAN AIRLINE MARKETS

Keith Acheson and Don McFetridge*

June 1991

1. INTRODUCTION

In recent years, the structure of the Canadian airline industry has undergone important changes. Air Canada (AC) has been transformed from a government-owned airline with special responsibilities and privileges to a private company competing on an even footing with Canadian Airlines International Limited (CAIL). Like AC, CAIL is a product of corporate metamorphosis and privatization. Pacific Western Airlines, formerly an Alberta Crown corporation, bought CP Air and Wardair to form CAIL. AC and CAIL link the larger Canadian centres and provide international service.

While these structural changes were taking place, regulatory change was also occurring. Elements of former policy were abandoned including the regional airline policy, introduced in 1966. In its stead, new and distinct regulatory regimes govern the North and the South. Although there is more regulatory flexibility everywhere in the system than previously, control over fares and route adjustment is tighter in the North than in the South. The regions and small communities are now served by feeder airlines, aligned by contract and equity positions with AC or CAIL, and a diminishing cast of independents. The feeders and independents provide commuter service

* Carleton University.

between smaller communities and link travellers to the networks of the major Canadian and foreign carriers. Some of the larger independents have offered service between large cities in competition with the majors.

The policy of restricting carriers to particular routes and of regulating fares,¹ frequency of flights and type of aircraft flown has either been abandoned or substantially relaxed. In the current regime, greater reliance is being placed on commercial incentives to determine flight characteristics and patterns. Federal regulatory attention has shifted to certifying ability to provide service; setting, monitoring and enforcing safety standards; providing certain airport facilities and establishing the terms of access to them.

Position papers, regulatory initiatives and legislation provide benchmarks in what was, and continues to be, an ongoing process of policy evolution. The important events, statements of intent and codifications of change in legislation include the pricing experiments initiated and monitored by the Air Traffic Committee of the Canadian Transport Commission in the late 1970s; the endorsement by the House of Commons Standing Committee on Transport of further regulatory liberalization in 1982; the announcement in 1984 of *The New Canadian Air Policy*, in which the Minister of Transport in the Liberal Government, Lloyd Axworthy, promised legislation to relax controls over routes and fares; the creation of the Canadian Aviation Safety Board in 1984; the release by the new Conservative Government in mid-1985 of the position paper, *Freedom to Move — a Framework for Transportation Reform*, and its review by the House of Commons Standing Committee on Transport;² and the passing of the *National Transportation Act*, 1987.

Experience under the new regime has confirmed some expectations and identified some areas of concern. As Soberman recently noted:

The euphoria that followed deregulation of the Canadian airline industry (influenced to a great extent by the U.S. Airline Deregulation Act of 1978) has now given rise to serious concerns about increasing fares, reduced competition, safety, and poorer quality of service, as a result of consolidation of all airline service into two national networks and the disappearance of Wardair.³

There is concern that the consolidation of the system into two dominant commercial networks has increased private economic power at the same time that reforms have weakened the ability of the government to respond

to that power. A number of questions are currently being posed. Will the forces leading to concentration leave Canada with only one major carrier? Is the mix of existing industry regulation and competition policy adequate to cope with the dominant duopoly or monopoly? Can domestic industry performance be made more effective by the development of an integrated North American or international regime for air travel? What are the implications of greater reliance on commercial incentives for the safety of the system? Can airport pricing and construction plans be developed to relieve congestion at airports? Can the system supply appropriate service to small communities?

This paper addresses an issue that cuts across a number of these questions — whether reinstating old or devising new regulatory measures to curb the exercise of market power is desirable.

2. THE PROBLEM

There are two aspects of single-carrier service that merit attention in the current policy environment. The first is whether the service is economically provided. Does the current commercial/regulatory system efficiently respond to the private demands of Canadian travellers? More specifically does it:

- encourage an appropriate choice of quality of service;
- minimize costs of providing that service;
- adequately discipline the pricing of the carrier; and
- realize timely introduction of new techniques and organizational reforms?

The second concern is whether collective demands, regional development, environmental goals and congestion alleviation are being economically served. In this study, we focus on the cost, quality and pricing aspects, and discuss dynamic aspects and collective responsiveness only to the extent that they interact with the former set of issues.

3. THE APPROACH

We begin by describing current regulatory controls over airline pricing and quality. We then report on Canadian price and quality experience from 1978 to date and selectively refer to American events. In the next section we

discuss economic models, explaining pricing and quality decisions where more than one, but not a large number, of carriers serve the market. We then to consider the ability of potential competition to discipline the industry, as incumbents realize that entry will occur if market power is exercised. *A fortiori*, the contestability of the market is important when routes are supplied by only one carrier. The flip side of market contestability is the degree to which barriers to entry naturally exist or are created by incumbents.

After considering the influence of market forces, we shift to private and public institutional defences. We investigate the ability of private organizational responses and contracting to deal with inefficiencies arising from the pricing and quality decisions of a single or small number of suppliers. Government affects the efficacy of these private responses through contract law (which delimits the promises that are enforceable and defines private and public enforcement roles) and through competition policy (which further restricts the arrangements a business can effect). Governments also intervene directly into different transportation activities — bus, rail and trucking. Sector-specific regulation is tailored to the idiosyncrasies of the technology and the transacting environment of the industry.

Finally, we explore how sector-specific regulation of prices and quality has performed for other transportation modes available to travellers and assess the applicability of these techniques to the airline industry.

4. CURRENT ACCESS, RATE AND FARE CONTROLS

The *National Transportation Act, 1987* establishes two zones with different regulatory regimes, southern Canada and northern Canada.⁴ Any flight beginning or terminating in the northern sector is governed by northern rules.

SOUTHERN CANADA

Section 72(1) of the Act states that the National Transportation Agency “shall issue” a licence to any applicant who:

- is at least 75% Canadian owned;⁵
- holds a Transport Canada operating certificate; and

- has the prescribed liability insurance coverage.

Potential entrants are not required to demonstrate that their proposed service is required by “public convenience and necessity” as was the case before 1987.

Section 76 also states that a licensee must give the National Transportation Agency 120 days’ notice before discontinuing or reducing to less than one flight per week any service the licensee has offered once a week or more for a period of six months or more. Agency permission to discontinue a service is not required.

With respect to fares, section 80(1) of the Act states that if, on receipt of a complaint in writing, the National Transportation Agency finds that there is no other alternative effective, adequate and competitive transportation service and that the carrier (licensee) has imposed an unreasonable basic fare increase, the Agency may either disallow the basic fare increase or reduce the increase in the basic fare by such amounts and for such periods as the Agency deems reasonable. Where practicable, the Agency may direct the licensee to make refunds to persons deemed by the Agency to have been overcharged as a result of the fare increase.

Section 80 does not apply to confidential contracts which carriers may enter into with various customers.

NORTHERN CANADA

Section 72(2) of the *National Transportation Act, 1987* states that the Agency “shall issue” a licence to any applicant satisfying the three conditions required of entrants on southern routes (Canadian ownership, an operating certificate and liability insurance) *and* if the Agency is satisfied that the issuance of a licence would not lead to a significant decrease or instability in the level of service. The onus is on those objecting to new entry to demonstrate that it would jeopardize the quality of existing services (reverse onus test).

A licensee may discontinue or reduce the frequency of a northern service to less than once a week provided it gives 120 days’ notice. As is the case with southern services, this applies to any service that has operated with a frequency of at least once a week for a period of six months or more.

Section 80(2) states that if, on receipt of a complaint in writing, the Agency finds that a carrier (licensee) has either an unreasonable basic fare *level* or has imposed an unreasonable basic fare increase it may disallow the basic fare *increase* or direct the licensee to reduce the level of, or increase in, the basic fare by such amounts and for such periods as the Agency deems reasonable. Where practicable, the Agency may direct the licensee to make refunds to persons deemed by the Agency to have been overcharged as a result of the basic fare level or increase.

APPLICATION OF PROVISIONS

The annual reports of the National Transportation Agency refer to several investigations of air fares on northern routes. There have been no investigations regarding the reasonableness of fares on monopoly southern routes to date, that is, no section 80(1) investigations. There have been several investigations of fares on northern routes, that is, section 80(2) investigations. There have been two decisions under section 80(2), both relating to the reasonableness of fares between Winnipeg and Lynn Lake, Manitoba. Decision number 187-A-1990 stated that the fare complained of was a discount fare rather than an economy fare (a "basic fare") and was thus beyond the jurisdiction of the Agency. The other decision (number 133-A-1990) addressed the reasonableness of the increases in the basic fare between Winnipeg and Lynn Lake. It found the increase, which amounted to 100 percent over two years, to be "not unreasonable." The Agency gave the following reasons:

- The carrier (Calm Air) was in a loss position overall.
- The Winnipeg–Lynn Lake service also operated at a loss, and the losses had been increasing.

These increases were apparently the result of decreasing load factors due to the decline of the local economy and to improved road access.

The National Transportation Agency has conducted investigations of fares on other northern routes under both section 80(2) and section 59. Under section 59 the Agency can investigate whether or not rates charged for the carriage of goods by air, water, rail or pipeline are prejudicial to the public interest. Routes which have been investigated include Winnipeg–Gillam,

Manitoba, Winnipeg–Oxford House, Manitoba, Yellowknife–Holman Island, Northwest Territories and Yellowknife–Coppermine, Northwest Territories.

The investigation of the Yellowknife routes provides some indication of the criteria the Agency is likely to employ in assessing the reasonableness of fares or the fairness of freight rates under sections 80 and 59 respectively. The criteria used in that investigation included:

- the overall profitability and rate of return on investment (long-term debt plus equity) of the carrier;
- the profitability of a “service” or related set of routes and the rates of return earned on them. (Profitability is revenue less direct cost less allocated indirect cost. Rate of return is profit divided by allocated investment);
- the profitability of individual routes (estimated as the basic fare times the average number of passengers plus estimated freight revenue per flight less direct and indirect operating costs per flight);
- comparison of fares over equivalent distances with equivalent equipment, load factors, service frequency and regularity; and
- accounting for the effect of infrastructure on costs. An example would be the restrictions imposed by short runways.

It appears that the Agency regards unreasonable air fares under section 80 and unfair freight rates under section 59 in much the same manner. A fare or rate on a route that is earning a normal rate of return is unlikely to be deemed prejudicial to the public interest.

While Agency decisions on and investigations of northern routes provide some guidance as to how the Agency might deal with monopoly southern routes, the latter are likely to be more complex. Both joint costs and revenues are likely to be more difficult to allocate to a particular route. Moreover, if there is a contestability problem, southern monopoly routes are going to be characterized by excess profits. Consequently, the concept of a normal rate of return will have to be more tightly defined for southern routes than on northern routes where rates of return appear to have been quite low. (An exception is Winnipeg–Gillam, Manitoba which, in the Agency’s view, was yielding “a reasonable return on investment.”)

5. PRICING AND QUALITY EXPERIENCE

This section provides background for assessing the adequacy of current rate and entry regulations and outlines what has transpired in Canada as commercial forces have been given more play.

In 1978, the Canadian Transport Commission sanctioned low-price experiments which were maintained until 1982. In this same period, all capacity constraints were removed from CP Air, advanced booking charters were allowed, and the major carriers began to develop feeder airlines. As a result, passengers on discount fares rose from 14.5 percent of traffic in 1978 to 37 percent in 1982. A sequence of new discriminatory pricing strategies was introduced by the scheduled carriers.⁶

DISCOUNTING

Discriminatory fares can result in a more economical use of resources when there are economies of scale and/or fluctuations in demand. For this to be achieved, the differences between the rights acquired at different prices (called “fences” in the industry) have to separate the travelling population into two groups which together generate sufficiently high load factors and average revenues to make the flight remunerative. The intent of the different rates would obviously be frustrated if all travellers chose the low-price option. The target is to discriminate, provide service to the marginal traveller at the incremental cost, and cover average costs by “taxing” inframarginal purchases. Discrimination can occur along a number of dimensions: different rates by time of day, by amenities on the same flight, by time of booking, by length of stay, by characteristics of the passenger and by frequency or volume of purchases.

Airlines add to their fleets depending on the prospective return. Price discrimination affects the return realized on each plane and therefore influences investment decisions. For many routes, adding an additional plane represents a discrete and sizable change in capacity. Ignoring the often small marginal costs of adding another passenger to a flight, the situation is one in which public goods, from which a customer can be excluded unless a specified price is paid, are being provided competitively. There is controversy in the economic literature on whether such competition results in over provision or the correct provision.⁷

In reviewing the evidence generated by the pricing experiments, the Canadian Transport Commission concluded that discount fares with appropriate fences raised load factors and encouraged new business. Price differentials succeeded in generating more nighttime traffic, but were less effective in smoothing out demand cycles between seasons of the year. The experience during this period also revealed that amenity-related fences — baggage, meals, placing on aircraft — were less effective separating devices than trip-related fences — advance notice, length of stay, time of travel, number of stopovers.⁸

Charter traffic also responded to new opportunities and increased fivefold. In 1982, 9 percent of all domestic low-priced air traffic flew on charters. This niche was dominated by Wardair, which carried 76 percent of charter passengers in that year.

The experiments of this period were controlled and closely monitored initiatives. Subsequently, the carriers enjoyed even greater freedom to manage their load factors through discriminatory pricing initiatives. Alfred Kahn, an architect of airline deregulation in the United States, had predicted that “much of the [price] discrimination will tend to disappear”⁹ with the relaxation of regulatory controls. In 1988, he reported his surprise at “[t]he persistence — indeed, intensification — of price discrimination.”¹⁰ Like their American counterparts, Canadian carriers took full advantage of the scope for discriminatory pricing.

Price structures have accordingly become more complex. Bailey and Williams provided the following American example:

For example, in 1978 the tariff department at Delta had twenty-seven employees tracking competitors’ fares and adjusting Delta’s prices. By 1984 Delta’s staff had grown to 147 employees monitoring 70,000 fares offered by Delta and its competitors, with the goal of optimizing some 5,000 price changes a day.¹¹

Management has become more sophisticated at choosing the number of discounted fares to be offered per flight, their depth and the appropriate fences to maximize the revenue generated per flight. Alice Peung, the manager of AC’s revenue enhancement programs claims that load management can raise a carrier’s revenue by 4 percent.¹² Because of these practices, load factors have remained higher, *ceteris paribus*, than they were under more rigid pricing.

During the 1980s, the proportion of travellers flying on discount fares rose in Canada until it stabilized at a level of about two of every three travellers. Charter flights became less important on domestic routes as discount fares became more widely available.¹³ Scheduled economy and business fares rose more rapidly than the consumer price index in the 1980s.

Although business class fares have risen relative to other fares, business travellers benefit from denser schedules and from frequent flyer programs. Of increasing importance is the ability of large businesses to negotiate confidential contracts with airlines for special fares. For obvious reasons, information about the extent of such contracts and their terms is fragmentary. There are occasional disclosures such as this 1985 account from an American business magazine:

Last year, Delta Air Lines entered into a special deal with General Electric in which it guaranteed specific fare discounts in exchange for an up-front cash prepayment of \$1 million as well as a guaranteed minimum volume of business. Some regional or business travel-oriented airlines have programs in place that give companies free tickets or credit toward future ticket purchases based on total sales volume.¹⁴

The availability of discounts and their average depth also varied by length of flight and by the regulatory regime governing air travel. For short-haul flights, high load factors are important in maintaining viability, as the following account of the challenges facing a small commuter airline attests:

Financial success for a small independent regional airline like Skycraft, which flies 4,000 passengers a month, requires keeping the business loads up between 75 and 95 per cent and breaking into new markets where there is a lot of head-office-to-head-office travel.

In Skycraft's case, that means moving General Motors of Canada Ltd. employees from Oshawa to Windsor and Detroit.¹⁵

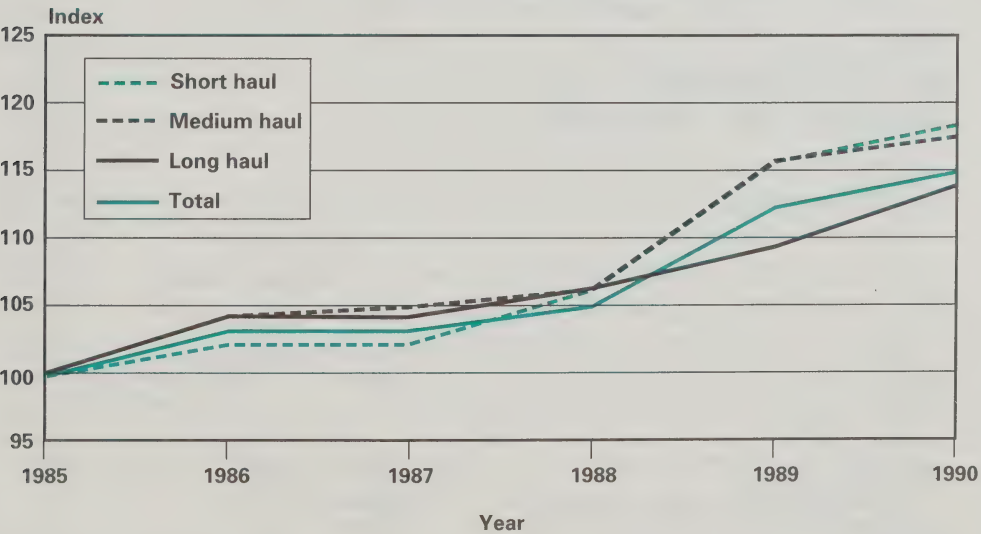
The scope for load management also depends on the stochastic flow and mix of customers with different reservation prices. Before deregulation, short-haul and less-travelled markets had lower load factors than average.

Despite the impact of low load factors on profit of short-haul routes, the relative number of seats available at discount are generally not as great on the short-haul flights offered by affiliates and independents.¹⁶

This reduced reliance on discounting indicates that discounting is a less effective instrument in generating higher loads for these routes. The difference in the pricing patterns by length of flight is reflected in the indices reported in Chart 1 for economy fares and in Chart 2 for discount fares.

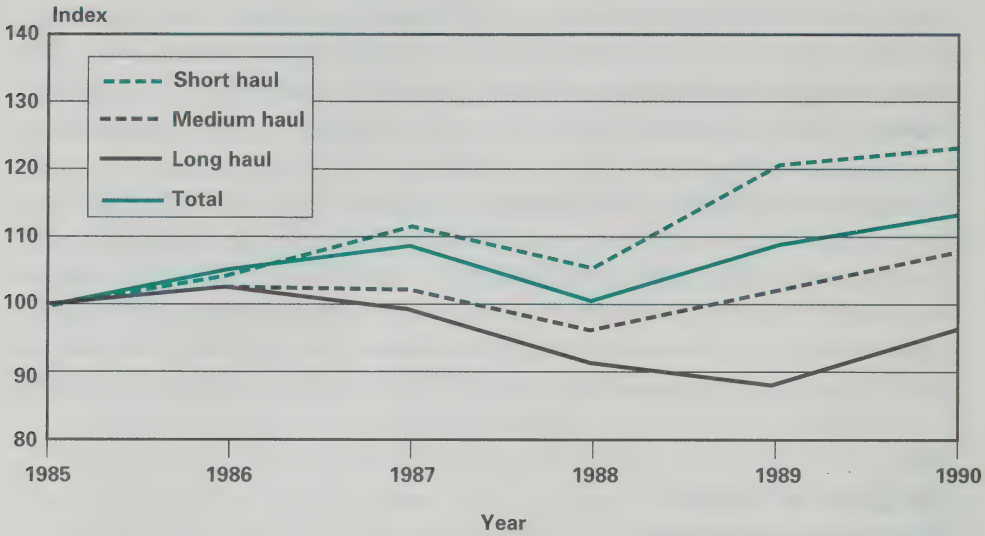
Statistics Canada reports that economy fares and discount fares grew less quickly in the more tightly controlled North than in the South from the fourth quarter of 1986 through to the fourth quarter of 1989.¹⁷ On the other hand, the level of discounts and their availability are lower in the North. The differences in economy and discount fare indices between the South and the North for the first quarter of the period 1985 to 1990 are graphed in Chart 3.

Chart 1
FIRST QUARTER AIR FARE INDICES
ECONOMY FARES BY LENGTH OF HAUL



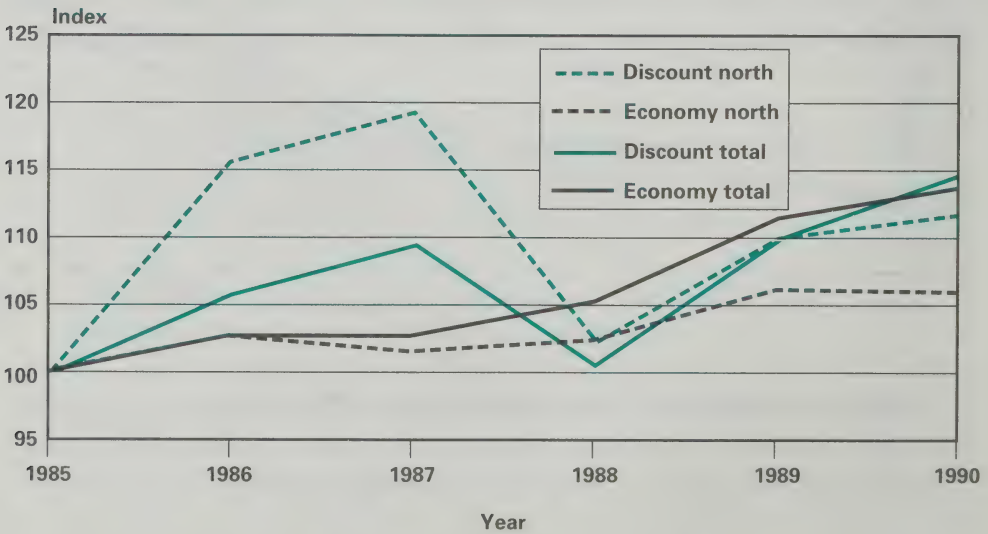
Source: Transport Canada, *Aviation Industry Review*, 1990, p. 35.

Chart 2
 FIRST QUARTER AIR FARE INDICES
 DISCOUNT FARES BY LENGTH OF HAUL



Source: Transport Canada, *Aviation Industry Review*, 1990, p. 35.

Chart 3
 FIRST QUARTER AIR FARE INDICES
 ECONOMY AND DISCOUNT FARES, NORTH/TOTAL



Source: Transport Canada, *Aviation Industry Review*, 1990, p. 36.

FREQUENT FLYER PROGRAMS

American Airlines introduced frequent flyer programs in 1981, and this volume discount scheme has become standard industry practice. According to one recent American survey, 72 percent of business travellers and 23 percent of leisure travellers now participate in at least one such program.¹⁸

The plans were introduced into Canada in 1984. Most of the members of the Canadian Commercial Flyers Association belong to two frequent flyer programs.¹⁹ A special appeal of affiliation with AC or CAIL is the right to participate in a program which can offer better and more varied rewards. Frequent flyer programs have been singled out by some analysts as a potentially potent barrier to entry.

TRAVEL AGENTS AND PRICES

A traveller commits a substantial amount of money and time to a plane trip. For the infrequent traveller much uncertainty and ignorance exist concerning the qualities and prices of service available. Travel agents represent an important instrument for overcoming the informational problems facing air travellers. Under the current price regime, the number of agencies and the percent of bookings done by agents have grown rapidly. As drug manufacturers often address their marketing efforts primarily to doctors, airlines focus on travel agents. In the United States, 81 percent of tickets on scheduled carriers were sold through travel agents in 1988. This compares to 56 percent before deregulation. At the same time, the number of agency locations rose from 14,800 in 1978 to 29,600 in 1987.²⁰ Customers' anticipation that information of benefit to them but not to the travel agent will be withheld grows as the fare structure becomes more complex and more tickets are sold through agencies. This problem has many dimensions and exists because of the commercial incentives for agents. Recent criticism has focussed on contracts between agents and airlines that provide for override commissions, premiums paid based on the volume of business transacted by the agent for the airline. Evidence is unavailable in Canada, but in the United States the importance of override commissions has increased dramatically. In 1986 over half of the travel agents surveyed received such commissions. Since it is estimated that travel agents are influential in dictating the choice of carrier in over one half of the non-business ticket and one quarter of the business

ticket sales, there is concern that agents will not direct the purchaser to the cheapest source but to the source from which they will receive the highest remuneration.

Large customers may generate sufficient business to warrant hiring services that audit their travel decisions and determine whether they have received good service or not.²¹ However, this defence is uneconomical for individual consumers who must rely on experience and the agent's reputation for probity. Competition among agents to create such a reputation may result in some agents voluntarily publicizing the general nature of their relations with airlines. Although a disclosure requirement might be helpful, the costs of consumer monitoring could be lowered by having travel schedule information more readily available. Ideally the information could be distributed over cable or telephone services such as the experimental "Alex" offering of Bell Canada. Terminals at libraries could extend the reach of the information. The *National Transportation Act, 1987*, section 83(1), requires that airlines make fare schedules available at their business offices, and a copy has to be made available to a customer on request at cost. The availability of discounts and their amounts and conditions change very quickly. A printed schedule distributed to a customer carries both irrelevant and outdated information.

COMPUTER RESERVATION SYSTEMS

Agents depend increasingly on computer reservation systems (CRS) to provide marketing information and travel services to their clientele. The potential ability of a CRS provider to bias this information for anti-competitive purposes has been a source of public concern. Court decisions and complaints to regulators²² prompted the promulgation of rules to govern the provision of CRSs in the United States. Display bias, pricing, contract length, terms and conditions of contracts and non-discriminatory access to enhancements such as direct access links are governed by the rules. In Canada the Competition Tribunal recently imposed a set of similar rules and constraints as a condition for its approval of the merger between the computer reservation systems of AC and CAIL.

OVERALL CANADIAN PRICE PERFORMANCE

The existence and importance of discount fares and the shift by the major carriers toward providing long-haul flights were contributing factors in making the revenue per passenger-kilometre for major Canadian airlines

grow at a much lower rate than the general price level. In the previous price structure, above-cost rates on long-haul flights cross subsidized rates on short-haul flights. As networks realigned in response to greater freedom of entry, exit and pricing, commercial pressures resulted in the fall of long-haul relative to short-haul rates. A further factor that reduced fares was the squeezing of economic rents earned by workers in the industry under the previous regulatory regime.

This process is well documented for the United States. Before deregulation, American labour relations in the airline industry were governed by the *Railway Labor Act*, which encouraged organization according to craft or class. Consequently, a fragmented bargaining environment had evolved. Thirty-four different unions had members in the American airline industry, many of which could halt activity at an airline. In 1958, the carriers responded by organizing the Mutual Aid Pact which provided for compensation of member carriers by other members for losses resulting from a strike.

Despite employer organization, wage rates in the industry were substantially higher than those for comparable jobs in other sectors, and many restrictive labour practices were adopted. For example, United's labour contract required it to fly its Boeing 737s with three flight crew members while Piedmont and Southwest flew with two. Thorncroft reported that: "By 1980, many pilots were flying only 44.3 hours per month, despite guaranteed payment for 75 hours per month and FAA regulations that permitted pilots to fly 100 hours per month."²³

The differences between the union wages under regulation and the market wages for similar skill levels were substantial. Moore noted that: "Nonunion pilots earn as little as \$32,500 per year for flying a Boeing 737, compared to the \$102,000 salary that United Airlines pays one individual to pilot a similar craft."²⁴

Labour represents the industry's largest operating expense. Between 1978 and 1986, 14 of the new entrants in the United States were non-unionized. Generally, these companies had fewer senior personnel, operated under less restrictive work rules and hired more part-time workers to meet peak demands. Labour costs were consequently much lower for the new entrants.

For example, in 1984 USAir's average pay and benefit package per employee was \$47,896, while that of People Express was \$17,139, providing People Express with a significant cost advantage. Similarly, when Continental underwent bankruptcy and was able to replace its unionized labor force with nonunion workers, its pay and benefit package was reduced from \$36,875 (per employee) in the first quarter of 1984 to \$23,433 by the fourth quarter of that year.²⁵

Many of the unionized airlines threatened to create non-unionized affiliated carriers, and some did. For example, Texas Air and Frontier Airlines formed New York Air and Frontier Horizons. Litigation and strike activity increased. In 1979, 4,075 person-years were lost, and 24,968 employees were involved in strikes in the United States.²⁶ The intensity of labour confrontation abated in the early 1980s. Two-tier labour contracts and beneficial changes in work rules were introduced. Under the two-tier contracts, new employees were paid considerably less than existing ones. For example, at American new pilots started at 50 percent of the old pay scale. Most of the two-tier contracts called for the merging of the two tiers within a prescribed period of time. By 1990, the labour cost differences of the immediate post-deregulation period had narrowed considerably.

Similar but less dramatic effects occurred in Canada with a time lag. Both major carriers have pared their workforces. In December of 1989, CAIL announced the elimination of 1,900 jobs of which 1,017 were former Wardair jobs.²⁷ In October of the following year, Air Canada cut 2,900 jobs²⁸ and CAIL announced further layoffs and schedule retrenchments.²⁹

Overall price performance is difficult to assess because the number of different types of flights and the multitude of rates on any one flight result in particularly wicked index number problems. Nevertheless, analysts familiar with the industry make general statements about price performance. For the most part, they agree that there was a substantial fall in average prices during the transition period to the new regime. The following statement by Bence is representative:

In 1985, 64 per cent of passengers travelled in Canada at reduced rates, the average reduction in relation to the base rate being over 50 per cent. Since each level of reduction is subject to a varying

number of restrictions (advance bookings, length of stay), this means that a much wider choice of rate/service combinations were available to passengers.³⁰

Price performance has not been as impressive recently, and there is concern that without the stimulus of new entry, with less scope to squeeze factor rents and a higher degree of concentration, future performance will not match that of the past.

AMERICAN PRICING EXPERIENCE

The United States has experienced similar but not identical price patterns. Prices and the proportion of travellers enjoying discount rates appear to be more volatile in the United States than in Canada. In some years, the proportion of Americans travelling on discount fares has been considerably higher than in Canada. For example, Kahn notes that "90 percent of all passengers in 1986 travelled on discount tickets, at an average 61 percent below coach fare."³¹ The percent in Southern Canada was just below 60 percent.³² In the United States, price competition was vigorous in 1986, and financial performance was dismal.

Meyer and Oster provided the following overall summary of American experience:

Deregulation's effect on fares has varied with the type of ticket and the size of the market. First-class fares are relatively higher in most markets, although increases generally have been small (on the order of 2 percent). Unrestricted coach fares have decreased slightly in the largest and most dense markets, while average coach fares in medium and small markets have risen about 3 and 6 percent, respectively. More important, since deregulation discount fares have become available in over 80 percent of all markets, and well over 60 percent of all passengers were flying on discount fares by 1984, with these discounts ranging upward to 50 percent or more. Prior to 1976 discount fares were rarely available outside the largest markets and were seldom deeper than a 20 percent reduction. Discounting also has increased pressures to reduce standard coach fares in the largest markets.³³

As in Canada, recent price performance has not been as impressive as in the transition period.

PRICES AND NUMBER OF CARRIERS

Statistical exploration of the American evidence of the connection between the number of suppliers and the exercise of market power in pricing, indicates that fewer suppliers mean higher prices. The studies differ in quantifying the effect of actual competition and in assessing the impact of each additional competitor. For example, Moore reported:

The evidence is impressive: those markets served by only one or two carriers experienced price jumps of over 40 percent in real terms, while those markets served by five or more carriers enjoyed fare increases of less than 3 percent. The ratio of fares does decline slightly from one to four, but the big drop is with the five-or-more-carrier market.³⁴

Morrison and Winston, working with a different period and specifications, concluded that if suppliers fell from two carriers to one, an "average round-trip fare on the route would increase about 9 cents a mile, or \$89 when evaluated at the mean sample distance of 983 miles."³⁵ In contrast to Moore, they found that moving from three to two competitors had much less effect and the bulk of the benefits derived from having multiple suppliers is realized by a duopoly.

NETWORK RECONFIGURATIONS, AFFILIATES AND INDEPENDENTS

As well as changes in fare levels and structures, there were important changes in the many dimensions of service — frequency of flights, reliability of scheduled information, air time, ground time, baggage service, ticketing, registration, waiting room comfort and convenience, and flight amenities — which affect the quality of air travel. Under the more permissive environment, mergers and alliances have resulted in larger organizations coordinating equipment, standards and flights. The resulting networks provide a different quality of service than formerly.

A characteristic of the new network is the hub-and-spoke configuration. Flights are scheduled to reach and leave a hub airport during a common

period of the day so passengers originating at different centres can be pooled to proceed to other destinations.³⁶ However, some passengers may be worse off if an indirect flight through a hub replaces a direct connection. In the United States, the large airlines have a number of main hubs and regional hubs. Delta, for example, has four anchors (Atlanta, Cincinnati, Dallas and Salt Lake City) and four corner posts or supplemental hubs (Boston, Los Angeles, Orlando and Portland).³⁷ In Canada, which has a belt rather than a grid as a geographic market, the economies realized from a hub-and-spoke configuration are less than in the United States. Nevertheless substantial realignments have occurred. Toronto has become a major hub,³⁸ and Montreal and Vancouver are regional ones. Unlike the United States, there are few connections in Canada which would be competitively served by more than one hub-and-spoke system, that is, two systems connecting destinations through a different hub.

As flight patterns were being realigned, AC and CAIL concentrated on the longer flights on dense routes and established feeder affiliates either directly or through acquisition or contract. Affiliation agreements typically coordinate timetables and integrate the local airline into the baggage, reservation and frequent flyer systems of the major carrier. The structure of the affiliate groupings frequently changes. The components of the AC and CAIL families in 1989 are presented in tables 1 and 2. The general strategy of AC is to control its affiliates while CAIL typically owns less than 50 percent of its partner companies.

Table 1
 AIR CANADA CONNECTORS

Carrier	Linkage	Fleet
Air Nova	AC owns 49%	4 jets, 9 non-jets
Air Alliance	AC owns 75%	9 non-jets
Air Ontario	AC owns 75%	17 non-jets
Air Toronto	Code sharing, AC agreement to purchase	10 non-jets
Air BC	AC owns 85%	5 jets, 26 non-jets
NWT Air	AC owns 90%	2 jets, 6 non-jets

Table 2

CAIL CONNECTORS (CANADIAN PARTNERS)

Carrier	Linkage	Fleet
Canadian North	Division of CAIL	8 jets
Air Atlantic	CAIL owns 45%	15 non-jets
Ontario Express	CAIL owns 49%	19 non-jets
Frontier Air	Ontario Express owns 100%	11 non-jets
Calm Air	CAIL owns 45%	15 non-jets
Time Air	CAIL owns 46%	3 jets, 28 non-jets

The affiliates fly smaller aircraft and serve commuter traffic as well as providing connector flights.

In addition to the affiliates, independents and small commuter airlines provide scheduled service. In 1987 there were 103 such carriers. Another 634 offered commuter charter service.³⁹ Many of the new independent entrants have not survived, despite success in their early years. In 1985 the President of City Express stated:

City Express either would not exist or would have taken years to reach the point we are at today had it not been for this new Canadian Air Policy. It gave us the chance to (achieve) flexibility in pricing and the ability to have a simple fare structure with no confining conditions of sale. The skies are opening up at last.⁴⁰

By July 1990, the company was in financial difficulty with four of its Dash 8 aircraft seized by Mutual Life of Canada. Formal bankruptcy followed. Wardair, another early enthusiast for and beneficiary of deregulation, ran into financial difficulties and was taken over by CAIL in early 1989. With Wardair's departure, Intair, a former partner in the CAIL system, became the largest independent. Intair challenged AC and CAIL on the dense Montreal-Toronto-Ottawa axis. By March of 1991, Intair was in financial difficulties and was reported to be negotiating to return to the CAIL fold.

SCALE AND NETWORK ECONOMIES

The misplaced enthusiasm of the executives of Wardair and City Express over the opportunities created by deregulation was reflected by academic opinion which generally underestimated the significance of scale and

network economies in air travel. For example, Gillen, Oum and Tretheway claimed:

A small network carrier should not have a *cost* disadvantage, provided it achieves traffic densities within its small network similar to those of Air Canada and CAI. Two independent carriers for whom data was not available, Wardair and City Express, appear to be attempting to build such small, but high density, networks in scheduled markets.⁴¹

By the date of publication of their article, Wardair had disappeared, and City Express was a few months away from bankruptcy.

Caves, Christensen and Tretheway introduced a distinction between economies of density and scale.⁴² Earlier empirical work had shown substantial unexploited system scale economies for the U.S. local airlines. This seemed to be confirmed by the higher unit costs experienced by the local carriers. Caves et al. were persuaded to make the distinction by the paradox that despite these cost differences the local airlines were competing effectively with the trunk carriers and gaining market share from them. Their study covers American airlines between 1970 and 1981. The Gillen, Oum and Tretheway study adopted similar categories and distinctions in analyzing the Canadian data from 1964 to 1981.

Caves et al. distinguished between the size of the networks and the transportation services provided within them. Output was captured by four variables: volume of services provided, load factors, average stage length and a network variable. Economies of scale occur if the increase in total costs is less than proportional when an increase in the size of the network and services provided takes place (keeping load factor and average stage length constant). Density economies occur if unit costs fall with an increase in services within a network of a given size (keeping load and stage length the same). The size of a network is measured by the number of points served. For example, a network linking Toronto, Hamilton and St. Catharines would be the same size as one linking London, Toronto and Los Angeles. Average stage length catches the effects of distance between points in the network. Gillen et al. used a hedonic measure of output and a similar separate measure for the network effect.

Caves et al. identified the scale elasticity as 1.07. As a result, they did not reject the notion that the "true" value is 1.00 and then concluded that there are no economies of scale. In this context, there is no reason to maintain constant returns-to-scale as a null hypothesis. It would be more instructive to report a confidence interval for the scale measure. In any case, the most likely value of returns-to-scale is 1.07. Larger American airlines would then be expected to have lower costs. In comparison, for Canada, Gillen et al. found that the best estimate of returns-to-scale for 1980 was 0.881 for Air Canada and 0.96 for CP Air. Based on their variable cost function estimate, an overall returns-to-scale of 0.992 was reported, indicating that larger-scale airlines face a cost disadvantage in Canada.

William Jordan came to the same conclusion with respect to the importance of scale and density. In an article contrasting American and Canadian events during the period 1978 and 1984, when American regulation was more relaxed than Canadian, he noted:

This clear contrast between the Canadian and U.S. experience provides unambiguous evidence that one effect of deregulation has been to increase both the number of jet airlines entering the industry and the number in existence at any point in time.⁴³

The gap between the econometrically measured scale effects in Canada and the commercial lack of success of small carriers may be due to the difficulties of uncovering the details of an industry's cost relations from the data available with even the most sophisticated of econometric analyses. Gillen, Oum and Tretheway used the number of points served by an airline as a proxy for its degree of networking. However, the data for their analysis cover the period from 1964 to 1981 when routes, frequency of flight and type of aircraft flown were regulated in Canada. The authors faced the difficult task of inferring the degree of network economies from a crude proxy and data generated by a situation in which the companies were constrained in choosing route structures. The subsequent restructuring in Canada and the American experience suggest that the regulatory constraint was binding. Graham and Kaplan noted that in the United States the Civil Aviation Board "often retarded the development of a highly integrated route network."⁴⁴

Based on the Canadian evidence from this earlier period, one might conclude that any airline could add service between a particular city-pair at a similar or lower cost, regardless of the exact configuration of its existing network

and where it was located in relation to the city-pair under consideration. Indeed the smaller the airline the more effective service it would be able to offer.

This inference is doubtful. Sometimes demand characteristics, which outweigh the identified cost conditions, are posited as the reason for the advantages of size. This claim can be confusing.⁴⁵ If passengers are willing to pay more to deal with a large established airline, presumably it is because they anticipate more or better outputs such as baggage handling, availability of alternatives, care and information with respect to delays and security of service. These services are part of an airline's output. Their exclusion in output measures in the econometric studies represents a specification error.

The paradox of successful competition from the local airlines with the trunks, mentioned by Caves et al., and the crowded field of contesting carriers, observed by Jordan, was a temporary phenomenon. The present situation could not be described as one in which local airlines were competing effectively with the trunk carriers and gaining market share. Bankruptcy and merger have been more frequent experiences than expanding market shares.⁴⁶ Given the ubiquitous consolidation trend we suspect that cost economies are playing some role in the process, despite the econometric evidence to the contrary. Canadian analysts were not alone in underestimating the importance of scale in providing network airline service. In his candid admissions of what he had not anticipated about American deregulation, Alfred Kahn confessed in 1988 that:

We advocates of deregulation were misled by the apparent lack of evidence of economies of scale — the principal explanation of the differences in cost among the carriers appeared to be differences in their route structures, which we hoped to eliminate by permitting totally free entry and exit — and by the physical mobility of aircraft, which caused us to underestimate the other obstacles to entry.⁴⁷

CONGESTION AND ACCESS TO AIRPORTS

The schedules which take advantage of hub-and-spoke configurations have accentuated the traditional concentration of flights at peak travel times and shifted patterns of travel. As a result, congestion has occurred at some airports. In the United States, events associated with the 1981 air traffic

controllers strike made this problem more acute by reducing the number of flights that could be handled at airports.⁴⁸ A number of American airports adopted rationing systems during the early 1980s; today four major American airports⁴⁹ still have a slot control system in place.

Toronto and Vancouver airports have also experienced congestion. Toronto's Pearson airport experienced extensive delays and cancellations at the end of 1988.⁵⁰ The director of airport terminals at Pearson, Jim Mattick, stated: "I have pictures from Christmas [of 1988] showing 37 aircraft lined up for take-off. There were ground delays of 180 to 200 minutes."⁵¹

As a result, Pearson introduced a slot allocation system administered by committees of airport, airline and federal representatives. These committees consider international flight commitments, gate assignments, customs and immigration requirements and security arrangements in making their decisions.

In the United States, the carrier committees were originally granted antitrust immunity with respect to their allocation of slots. As a result of concern within the Federal Aviation Administration (FAA) that the committees were biased against new entrants, a more market-oriented system was introduced in 1985. A block of slots at the four constrained airports were assigned to incumbents, who could then buy or sell them.⁵² The FAA can take a slot away from a carrier, and there is a "use it or lose it"⁵³ provision. Within those slots allocated to carriers, a subset are designated for international use. Transfer of these international slots is further restricted.

A new entrant can purchase or lease a slot from a current holder or acquire one through the lottery used to allocate new slots and those returned under the "use it or lose it" rule. Estimates of the value of a slot vary from US\$800,000 to US\$2,000,000.⁵⁴ The slot market is not characterized by the law of one price. In 1990, Secretary of Transportation Skinner's task force reported: "The slot holder must be concerned about the uses to which another operator will put the slots, regardless of whether the potential buyer is a new entrant to the specific slot market or an existing carrier in that market."⁵⁵

America West has complained that carriers selling slots have not notified all potential buyers of the availability of the slot. It cites a sale by Alaska Airlines of a slot at O'Hare to United as a case in point.⁵⁶

The Americans have favoured entrants or small incumbents in allocating new and returned slots. In 1980 the Secretary of Transportation ordered that 28 slots at Washington National be given to five new carriers. More recently, 25 percent of new and returned slots allocated by lottery have been reserved for new carriers. Windfall gains have accrued to new carriers which in many cases have not used the slots in their own operations, but have sold them. Of the 145 slots made available to new entrants or to limited incumbents (those with less than eight slots) only 15 were in use by the carriers that first received them. Some slots were returned to the FAA without use, but the majority of them were sold or transferred in a merger.⁵⁷ In Canada, there are no special provisions for new entrants. Recently, Intair had what the National Transportation Agency of Canada termed a "well-publicized skirmish with Canadian Airlines over landing-slots at Pearson International."⁵⁸

As well as access to the runway, an airline requires rights to gates, waiting room space, ticket counters, airplane service facilities and luggage pickup areas. In the past, American airport ownership patterns and financing arrangements have been more varied than in Canada where major airports have been largely owned, financed and operated by the federal government. Recently, Canadian policy has encouraged more private participation in financing airport expansion and the transfer of ownership of existing facilities to local interests. Reaching agreement with local interests affected by the expansion of airports and the development of new runways and air traffic control techniques has been difficult in both Canada and the United States. Sometimes incumbent airlines have been covert partners with affected residents in delaying measures to address airport congestion.

Prices for slots, terminal space and other airport facilities provide information about scarcity and aid in making informed investment decisions. Private ownership of the rights encourages the evolution of pricing schemes that reflect the time-of-day patterns of scarcity. Rental prices charged by a private owner for use at prescribed times would reflect the value realized by the user. More crudely, airlines could adjust their positions by swapping access rights under their control. With public ownership, the authority could charge time-of-day user-costs to carriers. The gains from obtaining better information and coordination are substantial. Morrison and Winston estimated that optimal airport pricing and investment would generate benefits of \$11.0 billion in the United States.⁵⁹

Adoption of such measures has been slowed by the political opposition of those adversely affected. As it has with privatization schemes, distributing the private rights judiciously among those affected may broaden the base of support, as the recipients gain a valuable resource. As well as encouraging privatization of airports, Transport Canada has announced a target of having users finance the airport services it provides. Charges for landing, terminal use, aircraft parking and fuel have been raised, and plans to introduce general terminal fees at airports have been announced. Understandably, the industry does not welcome paying for scarce resources it was previously receiving on concessionary terms. The Air Transport Association of Canada's chief executive officer claimed that a full recovery of costs by Transport Canada on regional service will mean an additional \$300 million charge.⁶⁰

FLIGHT QUALITY

When flights are switched from a major carrier to one of its affiliates or an independent, jet service is frequently replaced by propeller-driven service and smaller aircraft. Comfort generally decreases while scheduled frequency and duration of flight increase. Some additional direct flights are introduced, and some existing ones are replaced by connections through a hub. In other cases, indirect routing through a hub provides a competitive alternative to the direct flight. For example, Air Toronto provides city-to-city service on 152 flights per week from Toronto to eight eastern US destinations. This Air Canada affiliate does not fly into major American airports. A traveller could fly directly between Toronto and Indianapolis on a propeller-driven plane or fly there by jet, transferring in Chicago.

How all these changes affect the quality of air travel on shorter flights depends on the traveller and the travel itinerary. The National Transportation Agency, the industry overseer, views the net effect on quality as positive:

When the major airlines withdraw from a market, the community usually receives improved air service in the form of increased direct flights and available seats from the regional affiliates. In some cases, competing air service is provided by independent carriers. Also, while the replacement services are usually provided with non-jet aircraft, some markets continue to receive jet service.⁶¹

On longer domestic flights, Wardair provided a different class of cabin and ancillary service than that offered by the traditional companies. Similarly, in the United States, a broad range of service qualities was offered during the early years following deregulation. At the low end of the market, People Express unbundled the pricing of different cabin services and provided many innovations in vying to become the McDonald's of the industry, as the following tribute from Moore described:

Started in April 1981 with three used Boeing 737s flying from its Newark hub to Buffalo, Columbus and Norfolk, People Express has grown in three years to serve twenty-two cities with forty-six planes and to employ over 2,300 people. Originally, it served only smaller cities in the East Coast area; recently, its network has expanded to include Houston, West Palm Beach, London, Los Angeles, Chicago, Minneapolis, and Oakland, where it offers the lowest-cost service available. Its success is based on offering low-cost air travel while requiring passengers to pay for all additional service. Thus if a passenger checks luggage, it costs \$3.00 per bag. Coffee and soft drinks are sold for fifty cents. Passengers pay for their flights after boarding the aircraft. No hot food is served on any flights, including the ones to London; even the premium-class passengers must pay for their meals. However, People's fares are the lowest to be found.⁶²

By the time Moore's article was published in 1986, People Express had gone bankrupt. The cabin and ancillary service options offered on long-haul flights today look remarkably like those that were in existence before deregulation and do not appear to be markedly more varied. The innovations of the deregulation period that pass the survivor test are hub-and-spoke networking and load management.

CHARTERS IN THE NEW SYSTEM

In Canada, a number of jet charter carriers have also entered the industry offering an alternative service to scheduled international flights. Recently, however, the exit of charter carriers has been more frequent than the entry. Of the 11 jet charter companies offering flights at the end of 1988, seven have disappeared. In 1989, Holidair and Minerve ceased operation and, in 1990, the exodus accelerated with the financial grounding of Vacationair,

Points of Call, Crownair, Odyssey International and Worldways. Consumers have suffered inconvenience and sometimes financial losses from these failures. As Oum et al. noted:

The recent bankruptcies of several Canadian charter airlines has resulted in substantial losses to consumers as their prepaid tickets were not protected in trust funds. To some extent this is a failure of enforcement rather than policy, although the existing law leaves it to provinces to set policies and many of them have not.⁶³

Ontario is one province with consumer protection legislation. However, the Ontario travel industry compensation fund does not become operative unless a tour operator has gone bankrupt, and therefore does not offer relief in situations like the Worldways collapse which affected over 8,000 passengers. Many received assistance through tour operators such as Carousel Holidays, Fiesta Holidays and Conquest Holidays, which contracted with the remaining charter companies to supply replacement seats.⁶⁴

SCHEDULED FLIGHT FREQUENCY

Travellers have more flexibility in planning and adjusting to unforeseen events when there are more flight offerings. Under the new regulatory regime, the frequency of flights in North America increased significantly. In Canada the number of cities linked by flights rose by almost 60 percent between 1983 and 1989.⁶⁵ In addition, capacity on major city-pairs has risen markedly.⁶⁶ Although we are prepared to attribute this rise to the greater pricing, exit and entry freedom, the relation may not be strictly monotonic. While both the South and the North of Canada operate under more flexible regimes than before, there are more controls in the North. Between 1984 and 1989 departure frequency almost doubled in the South but grew even faster in the North.⁶⁷ In the United States, traffic increased since deregulation by 50 percent (passengers) and 27 percent (departures) according to the Federal Trade Commission.⁶⁸

CARRIER COMPETITION

The National Transportation Agency publishes and interprets much useful data in its annual reviews. With respect to the degree of actual competition on Canadian routes, it noted that for 1989:

Among the top 146 domestic city-pairs, which account for almost 90 per cent of passenger traffic in Canada, the proportion served by two or more competing carriers rose from 44 per cent to 77 per cent between 1983 and 1989 . . . [and] . . . Among the top 42 communities in the network, where about 95 per cent of domestic passengers originate or terminate their travel, all had at least two scheduled carriers in 1989 while one-half of them were served by five or more airlines.⁶⁹

The accompanying exhibits indicate that among the top 146 city-pairs the proportion served by one or no carrier has fallen from over 50 to about 20 percent. The Agency has maintained in both its annual reports that there is no lack of competition in Canadian air service. In 1988 it stated:

Any appearance of diminishing competition in this situation is misleading. There is extensive duplication of route coverage between the Air Canada and Canadian networks, and regional affiliates face an increasing number of small commuter carriers as well as each other in the scramble for local or feeder traffic.⁷⁰

In its 1989 review the Agency commented:

Canada's airline industry has always been concentrated, but when examined closely, be it in terms of passengers carried, passenger-kilometres, arriving/departing flights, seats offered etc., it is less dominated by one airline now than it was before the relaxing of regulatory controls in 1984.⁷¹

Judging the effectiveness of competition is not an exact science. Although aspects of what is occurring support the Agency's conclusion, there are reasons for concern. First, after an initial period of entry the trend has been to consolidation. Some analysts feel that trend has not run its course, and in the future there will be one airline aligned by strategic alliances with foreign megacarriers. Bence provides a clear outline of this script:

First of all, the privatization of Air Canada lifts the political obstacles to a merger with Canadian International. In addition, the two carriers recently stepped up their cooperation in the area of computerized reservation systems through the joint creation of the Gemini system. . . . The second remark relates to the liberalization of trade with

the United States. Despite much reluctance and resistance, it is very probable that the total or partial creation of a North American air transport market is only a matter of time. . . . For some months, there has been a multiplication of commercial agreements between companies of different countries. Canada has not remained inactive in this area and Air Canada, for example, has concluded agreements with Air India and Singapore Airlines International. A new step toward the internationalization of the industry seems to have taken place recently with the exchange of assets between companies. In this regard, information has been circulating in the industry concerning a possible agreement between Canadian and Cathay Pacific Airways. The last phase of this evolution would be the creation of large multinational groups active on several continents.⁷²

Second, American evidence indicates that the demand price of a traveller is higher for a ticket with a carrier which offers more frequent flights on a route. The carrier with more frequent service experiences higher loads and can charge a higher price.⁷³ A small number of flights offered by other carriers may not exert substantive pressure on the pricing of the carrier offering concentrated service on the route. In documenting the existence of the relationship, Borenstein illustrated its strategic importance by quoting from a 1985 internal memo from an executive Vice-President to the President of USAir:

There is still much to do before we can be confident that we have established a northeast stronghold that is as impervious as possible. Ideally we should control a major portion of the traffic at each of the cities in the north east. The beauty of the niche strategy is not just the marketing identity and control that it gives us. In addition, it enables us to keep control of prices within our niche territory, thus insulating a significant portion of our traffic from the devastating effects of unbridled price competition.⁷⁴

Third, the amount of effective competition generated when a small number of carriers serves a route is moot. Will the competitors collude implicitly or explicitly, or will competition be intense? In the United States there has been concern with the rising concentration even though there is more overlap between different systems, and geography increases the number of alternative means of getting from one destination to another. Alfred Kahn,

for example, is not only concerned about the current level of concentration in the United States, but blames lax antitrust enforcement for the situation.

The concentration process reflected also what many of the advocates of deregulation would characterize as a lamentable failure of the administration to enforce the policies of the antitrust laws — to disallow a single merger or to press for divestiture of the computerized reservation systems or attack a single case of predation.⁷⁵

PROFITS

If concentration is a problem it should manifest itself either in increased profitability or in a costly dissipation of potential profits. The evidence fails to indicate that providing air service has become more profitable under the new regime either in the United States or in Canada. Nor is it evident that potential monopoly profits exist and are being masked or dissipated through slackness or some other avenue.

For the United States, Cunningham et al. summarized the evidence:

There have been a variety of studies of the profitability of the airline industry during its transition from regulation to deregulation. It is clear from these studies that most measures of profitability of the airline industry that are based on historical accounting returns have not improved since deregulation.⁷⁶

At first, the performance of the stock of new and old airlines in the United States reflected the optimism of many of the new entrants in the industry.⁷⁷ Since then the stock values of the established carriers has been mixed while that of the many entrants have either been bought out or have become worthless through bankruptcy. In 1989, four American airlines made profits of US\$1.6 billion while the five other major companies made a loss of US\$1.59 billion. The successful companies were American, Delta, United and Northwest. The only new entrant to survive, Midway, was then reported to be "low on cash and rumoured to be a merger candidate."⁷⁸ Since 1989, Midway has declared bankruptcy.

One concern of opponents of deregulation was that competition would be destructive, and carriers would not generate the profits that had been

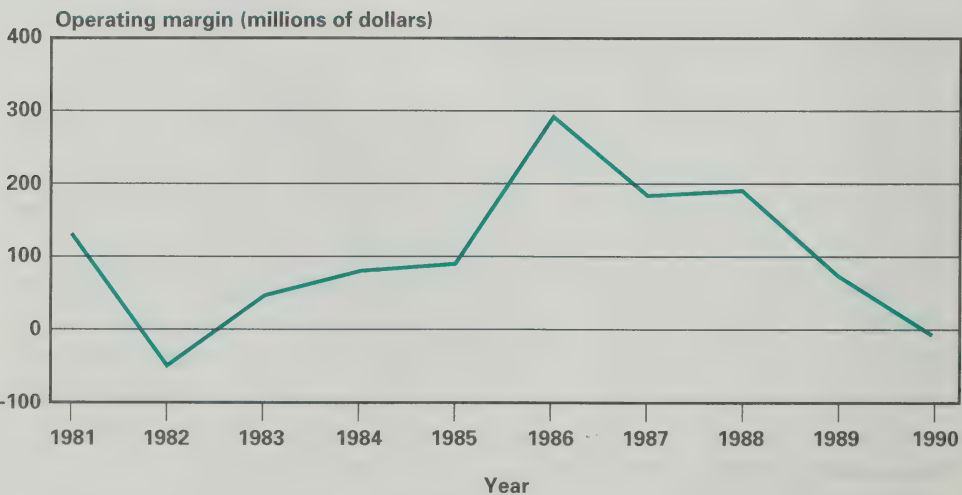
experienced with regulation and which were necessary for financing capital acquisitions.⁷⁹ Van Scyoc addressed that issue and concluded that:

It was not deregulation of airlines which adversely affected profits but, rather, the sluggish economy and rapidly rising fuel cost along with higher real interest rates. Deregulation has allowed the airlines to increase their average load factor (ALF), keeping their profits from falling even lower than they have.⁸⁰

The operating margin⁸¹ in Canada for Level I airlines⁸² is shown in Chart 4. The margin was negative in 1982 and 1990. Indices of the margin and two output measures are shown in Chart 5. As the index of passenger traffic and freight carried rises, the margin shows great volatility and no upward trend.

There has also been no decrease in the volatility of the margin over the different months of the year as shown in Chart 6. This volatility is consistent with greater stability in the load factor over the year. The industry is capital intensive, and interest costs can squeeze margins. The percentage of interest expenses to operating expenses is presented in Chart 7. There is

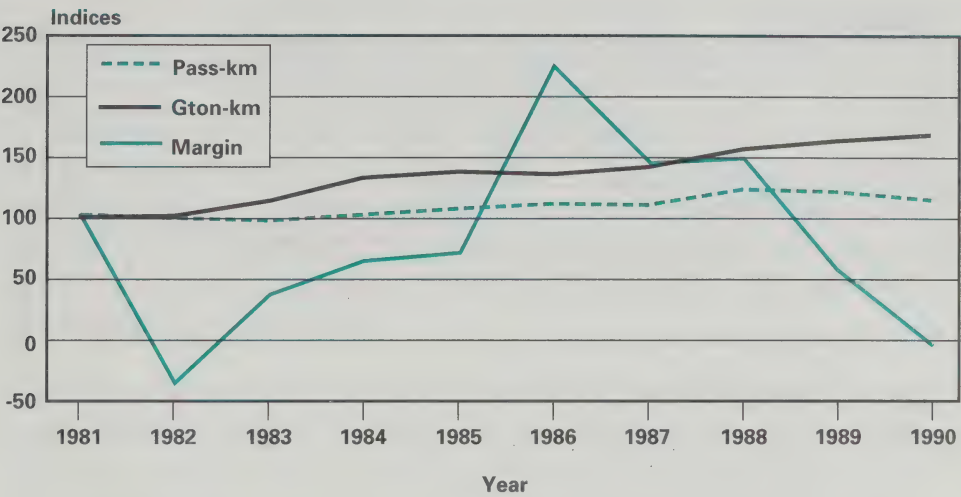
Chart 4
ANNUAL OPERATING MARGINS
LEVEL I CARRIERS



Source: Statistics Canada, CANSIM.

Notes: Margin is operating revenue (CANSIM D462215) less operating expense (D462216). 1990 margin estimated based on 9 months data.

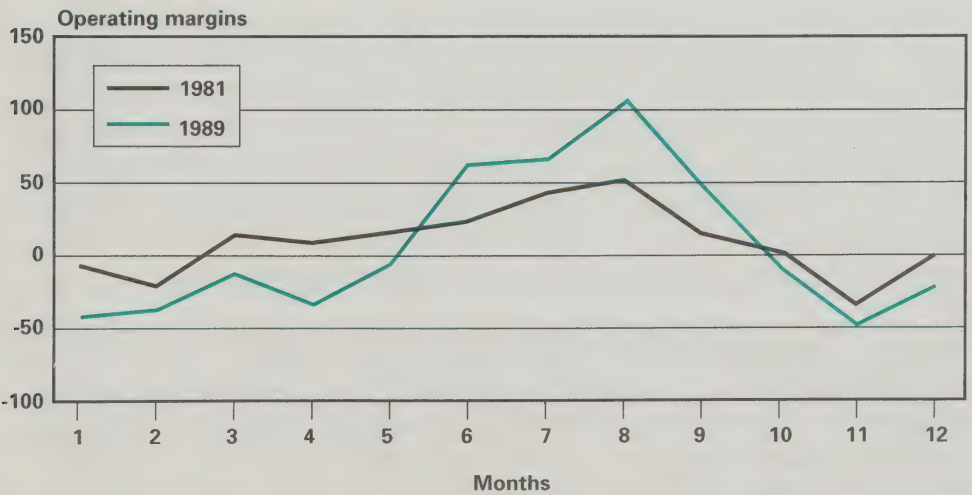
Chart 5
 AIRLINE LEVEL I, 1981-1990
 OUTPUTS AND MARGIN



Source: Statistics Canada, CANSIM.

Notes: Raw data from CANSIM D462210, D462212, D462215 and D462216.
Margin for 1990 estimated based on 9 months of data.

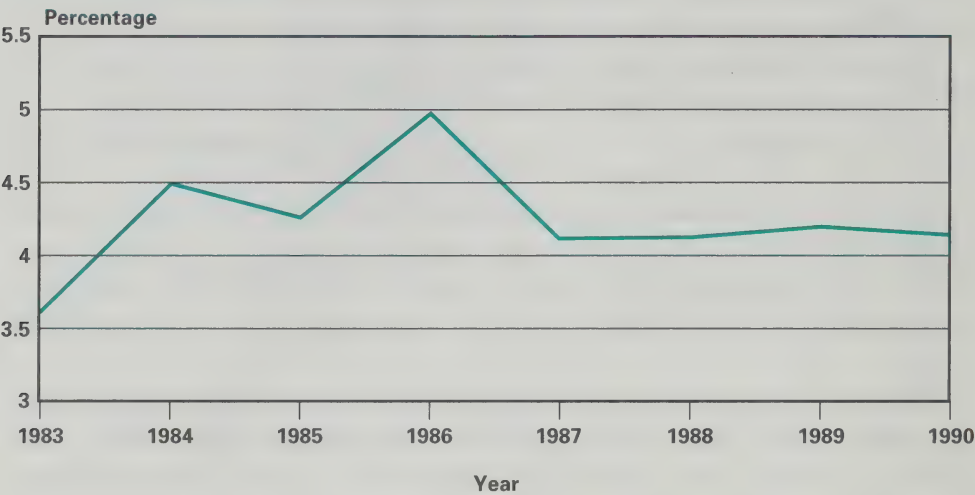
Chart 6
 MONTHLY PATTERN OF OPERATING MARGINS
 LEVEL I CARRIERS — 1981 AND 1989



Source: Statistics Canada, CANSIM.

Note: Margin is operating revenue (CANSIM D462215) less operating expense (D462216).

Chart 7
INTEREST TO OPERATING EXPENSE PERCENTAGE
LEVEL I CARRIERS — ANNUALLY



Source: Statistics Canada, CANSIM.

Notes: Interest (CANSIM D462217) to operating expense (D462216). 1990 based on 9 months.

no indication that this ratio has been extremely volatile, and it reached its peak in the same year, 1986, that the operating margin peaked.

Profit rates are hard to measure because of the difficulty of assessing economic depreciation in the industry, and reasonable observers disagree about their level. Despite this ambiguity, there does not appear to be any convincing evidence of excess profits, and no analyst claims the contrary. However, since the current regulatory regime has been in place for only a short time, the profit position of the airlines needs to be monitored closely. Considerable information is needed to identify trends in this industry, as profit margins are extremely sensitive to the general level of prosperity.

Another possible concern is that the lack of overall excess profits is masking monopoly profits on non-competitive routes. With no overall excess profits being made, compensating losses must exist on the other routes flown. If there are no scale effects, the airlines would presumably drop the losers, and excess profits on their total operations would surface. If instead, costs on different routes in a network are interrelated with each other and with

the details of the network configuration, profit margins on individual routes that ignore these interdependencies are misleading measures for guiding public policy.

6. THE NUMBER OF CARRIERS AND THE EXERCISE OF MARKET POWER

As noted above, the city-pairs accounting for the bulk of Canadian air traffic are served by two or more competing carriers. There remains some uncertainty regarding the number of carriers required on a route to ensure that passengers have the benefit of competitive pricing. Fares may be competitive with one carrier if a route is contestable. The question of contestability is addressed in the next section.

The relationship between pricing and the number of carriers presently serving a route depends on the nature of the interaction among them. If the flights offered by each carrier are indistinguishable, and each carrier sets its fares on the assumption that the other(s) will maintain theirs at existing levels, then fare levels will be driven to marginal cost when there are two or more carriers on a route. This is what is known in price theory as undifferentiated Bertrand competition. When competition takes this form, fare revenue is not sufficient to cover fixed costs. For this reason undifferentiated Bertrand competition is sometimes called destructive competition, and it cannot prevail over the long run.

If the flights offered by each carrier are indistinguishable, and each carrier assumes that rivals will maintain their capacity on the route and adjust fares to maintain load factors, then fares will exceed marginal cost — approaching it asymptotically as the number of carriers on the route becomes very large. This is known as undifferentiated Cournot competition. Taken together with free entry, this form of competition results in fares that approach average cost.

Incumbent carriers may also agree tacitly to share the traffic on a route in roughly fixed proportions. In this case fares will remain at the monopoly level regardless of the number of carriers on a route. It may become increasingly difficult, however, to maintain a market-sharing arrangement as the number of carriers on a route increases.⁸³

The assumption that the respective flights of individual carriers on a route are indistinguishable does not seem very realistic. Individual flights may be

differentiated with respect to departure times and connections as well as the number of stops, the airport (Pearson or Toronto Island, for example), the type of equipment and cabin service. In this case it is more difficult to talk about “the price” because the offering differs. If the differentiation is assumed to be simple, involving only departure times, for example, it can be shown that the fare charged by each carrier declines as the number of carriers on a route increases.⁸⁴ Fares would tend to marginal cost only if allowed departure times were virtually adjacent, and the fixed costs of offering a flight were very low.

The empirical evidence on the relationship between the fares and the number of carriers on a route in the United States has been cited above. Some evidence suggests that two carriers are sufficient to generate competitive fares. This is consistent with the undifferentiated Bertrand model. Other evidence implies that it may take up to five carriers on a route before competitive results are achieved. This is consistent with a variety of models including the Cournot, market-sharing and differentiated product models.

The unprofitability of most airlines in recent years is consistent with either undifferentiated Bertrand (destructive) competition or with free entry. The continuing presence of price discrimination is not consistent with undifferentiated Bertrand competition. Moreover, the evidence cited in the next section implies that barriers to entry do exist. Thus it is difficult to characterize the nature of the oligopolistic rivalry among airlines with any precision on the basis of the evidence presently available.

The relationship between fares and the number of competitors on Canadian routes does not appear to have been subject to the same type of analysis. The National Transportation Agency has noted the salutary effect of competition on fare levels in several of its northern fare investigations.⁸⁵

7. CONTESTABILITY

Contestability draws attention to the disciplining power of potential competition — of the deterring effect of credible threats to enter. The concept has venerable roots, although it has recently been rediscovered, refined and given a patina of mathematical respectability. The central idea was clearly and subtly articulated by Schumpeter:

It is hardly necessary to point out that competition of the kind we now have in mind acts not only when in being but also when it is merely an ever-present threat. It disciplines before it attacks. The businessman feels himself to be in a competitive situation even if he is alone in his field or if, though not alone, he holds a position such that investigating government experts fail to see any effective competition between him and any other firms in the same or a neighbouring field and in consequence conclude that his talk, under examination, about his competitive sorrows is all make-believe. In many cases, though not in all, this will in the long run enforce behavior very similar to the perfectly competitive pattern.⁸⁶

For the purposes of this study, it is useful to distinguish between contestability with a single homogeneous service and with a number of differentiated services. Consider first a situation in which each flight is sold at the same price, and the market is supplied by a single carrier. Perfect contestability describes a polar case in which the threat of competition has a maximum effect. It requires that the potential entrant have the same costs of providing service as the incumbent, and some mechanism should exist for the potential entrant to contract with customers without cost. The entrant can offer better terms to passengers currently flying with the incumbent, and they can switch their business without cost. To keep its existing customers, the incumbent has to price services with no excess profit. If that did not occur, the entrant could decrease the price slightly and still be viable.⁸⁷

If travellers can be separated into different groups by differentiating service, perfect contestability ensures that no subscribers to a service class can do better by contracting with a potential entrant alone, or in conjunction with some of the other groups, or in a grand alliance with every group being currently served.⁸⁸ When price discrimination is possible, there may be a number of price packages for each service class that defend the incumbent against the challenger. Each choice within this set of sustainable break-even pricing options involves different contributions from each of the service segments.

The more contestable a market is the less the need for oversight and intrusive regulation. The degree of contestability depends on the costs of potential entrants and the barriers to entry. For a particular route, other active airlines obviously have expertise in providing service. However the network configuration of the potential challenger can raise its cost of providing service to the particular city-pair. There may be none or only a few carriers

which can integrate the route as effectively into their current network as the incumbent. Foreign airlines with effective cost configurations would widen North American air service if they were permitted to enter the market. In Canada, American airlines with adjacent networks would be particularly effective.

Barriers to entry are costs imposed on an entrant but not the incumbent. In considering contestability and air service, many analysts focus on the fact that the airplane is not considered sunk capital and is suitable for "hit-and-run" entry into a market. In classroom discussions the airlines become the archetypal example of a perfectly contestable industry. The real world has failed to confirm that hypothesis. There are many other valuable informational and structural assets involved in providing airline service which are sunk. Barriers to entry alleged to exist in airline service include control over airport slots, gates and ancillary facilities; frequent flyer programs; feeder line control and hub or route concentration effects; computer reservation systems and graduated travel agent commission schedules; and predatory pricing.

The airport congestion problem in Canada is an issue at Toronto and Vancouver. It and various externality problems can be alleviated by appropriate pricing of airport facilities. Halting steps are being taken in this direction. Unfortunately, as noted by Soberman, the media have focussed attention on less-pressing issues than price rationalization and the provision of, and access to, airport facilities:

The capacity of Toronto's Pearson International Airport, for example, is a real issue. It affects costs, aviation safety, and congestion for much of the remaining domestic airline network. By contrast, recent VIA Rail cuts were perceived by numerous elected officials and newspaper editorialists to create real problems of urban traffic congestion. In fact, in a place like Toronto, all VIA services accounted for about 200 commuters per day in comparison with the 1.7 million passengers per day carried by all other public transportation agencies.⁸⁹

With appropriate pricing, an incumbent would face the same cost of using airport facilities as would an entrant, and no barrier to entry would exist. The gains appear to be so considerable from proper pricing of airport facilities that it is difficult to understand the political hesitation in implementing appropriate systems. Political opposition, however, is ubiquitous. For the United States in 1989, Hahn and Krozner reported that a small private

plane could land for a US\$6 charge while a commercial jetliner paid between US\$90 and US\$200, although both caused the same congestion effects. They also noted that Congress limits the revenue-raising abilities of airports which receive federal financial aid and has imposed restrictions prohibiting the levying of "passenger facility charges."⁹⁰ Perhaps the key to reducing this opposition is to stress the externality control features of pricing as well as the more obvious alleviation of financing difficulties.

Gaining access to congested airports can be avoided by entrants flying into and out of other airports that serve the same or contiguous areas as the major airport. The lack of such alternatives for major Canadian routes prompted Lazar to comment in 1984:

There are only two sets of markets that have the traffic potential to support multi-frequency, end-to-end, turnaround routes such as those flown by People Express and Southwest. These markets are the Vancouver–Calgary–Edmonton and the Toronto–Ottawa–Montreal–Quebec City groupings. In none of these cities however, is there a close-in and/or relatively uncongested second airport that can accommodate jet aircraft.⁹¹

Alternative airports exist for commuter traffic on shorter routes, and since 1984 the capabilities at a number of these have been enhanced. Partly in response to congestion at Pearson International, traffic has risen at Oshawa, Hamilton (jet capacity), Buttonville and Toronto Island airports. The development of even more sophisticated facilities at these alternative sites may enhance competition even further, although the area air traffic control problem may be exacerbated.

Another aspect of the airport access problem is the domination of an airport by one or two carrier systems. Feeder traffic is critical to the operation of a hub-and-spoke system. Density on many feeder routes can often support only one supplier. Contractual links or ownership of these feeders has been cited as another barrier to entry:

Air Canada and CAIL have been successful in purchasing all of the feeder carriers of any importance in Canada. By preventing their feeder subsidiaries from signing interlining agreements with other carriers, or from putting in joint fares with other carriers, they are excluding new entrants from the trunk airline routes from important

segments of trunkline markets. Just prior to Wardair's demise, it announced that it was going to pay feeder airline fares, at great expense, in order to get access to an important segment of the scheduled airline market.⁹²

Exclusive contracts and refusal to supply are addressed by competition policy. This study did not uncover any idiosyncrasies which would dictate a special governance arrangement for competition practice in air service.

Tretheway referred to frequent flyer programs "as a powerful entry barrier."⁹³ For the United States, Winston and Morrison estimated that a frequent flyer program is valued at US\$32.01 per round trip of an average length over their sample.⁹⁴ A fledgling or local airline offering service between points A and B only cannot launch an effective self-contained plan. The frequency of use by a typical business passenger may not generate enough credits and for the passenger who does qualify for a prize, winning yet another trip to B may generate minimal excitement. This is true regardless of whether the airline has just begun business or has been offering service for a long period of time. The barrier is not one of entry but of size. The constraint is that the "in kind" volume bonus has to be attractive. A non-airline company which could offer attractive prizes from its existing services could take over a single route and immediately implement a competitive marketing plan. Prizes for volume purchasing could take a number of forms including waiving credit card fees, "free" vacations at some resort or free options on the purchase of an automobile, depending on the price cost margins in the portfolio of products and services sold by the parent company. "In kind" bonuses are ubiquitous in Canada, ranging from Canadian Tire script to obtaining three for the price of two. Presumably, airlines offer prizes in the form of flights rather than other "in kind" prizes, because it aids their load management.

An affiliate of one of the two major carriers may enter into a single-carrier route, or if the existing carrier is part of one system, an affiliate of a rival system may enter. In either case, the entrant would already be linked into a viable frequent flyer program, and no competitive problem would arise from this source. A small airline, which is not an affiliate, can negotiate inclusion in the plan of AC, CAIL or a foreign line. For example, when Intair departed the CAIL family to fly as an independent, it offered a frequent flyer program in conjunction with American Airlines and KLM.

Organizational innovation is an important competitive element in a small-numbers service industry. Pricing innovations have diffused quickly in this industry. Since a frequent flyer program represents a pricing initiative which is attractive to customers, it is not advantageous to prohibit such plans⁹⁵ or tax them.⁹⁶ For a low enough level of tax that would still give entrants an incentive to develop their own plan, entry would not be encouraged. For a level of tax at which an entrant without a plan could compete with an incumbent who maintained a taxed plan, entry would be subsidized. A large enough tax that would prevent anyone offering a plan would be the same as a prohibition.

Remedies which enhance competition by other means than proscribing or implicitly taxing those dimensions of service which one class of firms (in this case large firms) does better are preferred. The frequent flyer plans may be improved by disclosure requirements or even access requirements. However, the most effective pro-competitive step available is to broaden access to alternative sources of frequent flyer schemes through a North American, or preferably more widely based, air pact.

Morrison and Winston estimated the competitive advantage conferred by a hub to be US\$25.66 on their representative American round trip.⁹⁷ The new hub-and-spoke network architecture reduces costs and increases the options available to passengers. Canadian independent airlines can make feeder arrangements with American airlines but opportunities are restricted. Before its demise, City Express was negotiating a feeder agreement with Continental Airlines. Reducing the barriers to competition between configurations can be encouraged by removing artificial barriers to international competition. In this instance, a North American pact would be sufficient to harvest the benefits.

What is essentially a regulatory remedy has been adopted in Europe, the United States and Canada with respect to computer reservations systems.⁹⁸ In 1989 after submitting the proposed merger between Reservec and Pegasus to the Competition Tribunal, the Director of Investigation and Research amended his application to permit the merger if a proposed set of rules governing the operation of the enlarged computer reservation system (CRS) and access to it was adopted. The Tribunal approved the merger and a modified set of rules in mid-1989. After the decision, the Minister of Transport directed Transport Canada and the National Transportation Agency to develop appropriate regulations for reservation systems.

Even with a regulatory code in place, a barrier to entry may persist. Algorithm bias, halo effects such as commission payments based on CRS use and the market intelligence conveyed provide the CRS carrier with an alleged advantage. In evidence given before the Canadian Competition Tribunal, Michael Levine, a dean at Yale and a former executive at New York Air, argued for divestiture of CRSs from the carriers because regulation of them had failed:

Detailed regulation in the United States has succeeded in reducing incremental revenue effects from 50% or more to “only” 13–20% (estimate of the U.S. Department of Transportation), but that compares with typical airline profit margins of 5% or less, suggesting that even a heavily regulated CRS sufficiently distorts competition to make the difference between success and failure in the marketplace.⁹⁹

Although the same remedy was not recommended, the same concern was expressed by the American Secretary of Transportation’s Task Force:

In certain areas, the rules have not worked as well as the CAB originally intended. In particular, the issue of allegedly restrictive subscriber contract provisions and practices remains one of the most controversial CRS policy issues. In addition, the prohibition on display bias did not eliminate incremental revenues. Finally, the CAB’s expectation that booking fees could be controlled by the bargaining power of major non-vendor airlines proved unrealistic.¹⁰⁰

Bailey and Williams argued that the CRS and hub domination reinforce each other as barriers to small entrants:

One use of these (CRS) systems has been to enhance high-volume control at large hubs. Since deregulation, American’s system has come to process 88 percent of ticket sales in the Dallas–Ft. Worth market, United’s to process 72 percent of the Denver market, and United’s and American’s together to process 83 percent of the Chicago market. Thus, American’s Sabre System and United’s Apollo dominate the travel agencies in the hubs where these carriers operate, enhancing local monopoly rents from these geographic regions as well as scale-based rents over competing carriers. Both sources of rent are thereby enhanced for these especially powerful carriers.¹⁰¹

In Canada, entrants that have challenged AC and CAIL have subscribed to other CRSs. Intair aligned itself with Sabre, the CRS of American Airlines. Before its takeover by CAIL, Wardair subscribed to System One, the CRS of Texas Air.

In the United States major airlines which do not have their own CRS often buy into the CRS of another carrier. For example, the Apollo system is 50 percent owned by United, with Alitalia, British Airways, KLM, Swissair and USAir owning 49.9 percent. Air Canada recently purchased a 1 percent interest in this system. These private responses are similar to those observed with bottleneck facilities like pipelines. They provide some insulation to existing airlines from exploitation. They may make entry easier by setting a precedent on the terms and conditions for ownership participation in a CRS, but on the dark side they may facilitate the coordination of incumbents to fight entrants.

CRSs require common conventions as to format and structure. For this reason, as well as the barrier to entry concern, regulation may be warranted. However, regulation may affect system innovation. In the personal computer field there has been a spectacular increase in power, an explosion in sophisticated applications and a marked reduction in real price generated in the absence of regulated operating systems and architecture.

To what extent would the adoption of such systems have stifled this remarkably innovative sequence? Any stifling effect has to be balanced against an acknowledged reduction in a barrier to entry. The Bureau of Competition Policy and the Tribunal are not the best-suited institutions for ensuring a flexible evolution of the code which neither penalizes nor subsidizes entry.

Graduated travel commissions are another source of concern for small entrants. Although traditionally based on total ticket sales, more recently airlines have been targeting particular routes. The rate of override also increases as the share of the agency's business directed to the carrier increases. Large organizations can protect themselves by private contracting and these arrangements reduce the barrier to entry. An example is provided by the American Department of Transportation's Task Force:

Consequently, some agencies and corporate clients have shifted to a fee system, where the client, not the air carrier, compensates the agency by paying it a fee. The agency then turns its airline commissions

over to the client. The fee system would help ensure that the agency's incentives are consistent with the client's expectation that the agency's goal is to provide it with the best possible service and would reduce the uncertainty facing an agency over the adequacy of its compensation.¹⁰²

With respect to predatory pricing, it is alleged that carriers have selectively responded to incursions by low-price entrants or by other incumbents breaking an established pattern of pricing. As of the end of June 1990, the U.S. Justice Department was proceeding with four cases involving alleged collusion, predation or attempts to block entry against airlines.¹⁰³ In one of these cases carriers are accused of using the monitoring service of the Airline Tariff Publishing Co. to signal that new fares were in retaliation for the pricing aggressiveness of others.

People in the industry say the crucial signals between carriers translate as follows: "Let me determine the price at my hub airport and I'll let you do the same at yours." The most common — and perhaps most questionable — "discussion" between airlines is played out like this: Carrier A, often a smaller operator such as Midway Airlines or America West, attempts to boost its business by lowering ticket prices. It enters lower fares in the industry's computer system. In response, Carrier B, the dominant carrier at the affected airport, not only matches the new fares, but lowers them in other markets that are served by Carrier A.¹⁰⁴

Alfred Kahn did not include predation in his discussion of the events that surprised him about airline deregulation:

I take perverse satisfaction in having predicted the demise of price-cutting competitors like World and Capital Airways if we did nothing to limit the predictable geographically discriminatory response of the incumbent carriers to their entry, and in having rejected the conventional wisdom that predation would not pay because any attempt to raise fares after the departure of the price-cutting newcomers would elicit instantaneous competitive reentry.¹⁰⁵

Although competition policy can be mobilized if the threat of predation is a significant barrier to entry, there are particular difficulties in identifying predation in the airline industry. Prices are frequently regarded as predatory, if they are below incremental cost. Incremental cost is an ambiguous

concept in this industry. It depends on the time horizon adopted. It also may have an opportunity cost component that will not be evident in accounting data. Prices are sometimes regarded as predatory if they are discriminatory. In this market discrimination is likely to be a feature of an efficient equilibrium. It may be difficult to determine when a change in the degree of discrimination is a predatory rather than a competitive response.

8. INSTITUTIONAL DEFENCES AGAINST MARKET POWER

City-pair airline markets may not be perfectly contestable, and the lower density routes may be characterized by small-numbers competition. This raises concerns about the possible exercise of market power on these routes. Contestability may be limited by any or all of incumbent control over airport slots, gates and ancillary facilities; frequent flyer programs; feeder line control and hub or route concentration effects; computer reservation systems and graduated travel agent commission schedules; and predatory pricing.

Defences against the exercise of market power may take any of three forms:

- further deregulation;
- residual rate and fare regulation and/or the imposition of access requirements; or
- decentralized public and private responses.

FURTHER DEREGULATION

While it does constitute a considerable liberalization of the regulations governing air and other forms of transportation, the *National Transportation Act, 1987* retains some provisions which may reduce the contestability of city-pair airline markets. The first of these is the requirement under section 76 that an incumbent carrier give 120 days' notice before dropping a service or reducing its frequency to less than once a week. This requirement effectively commits the incumbent to the market for four months regardless of the attractiveness of a new entrant's offering. The legislation requires the incumbent to "stand and fight" and undermines the contestability of the market as a consequence.

A second restrictive feature of the Act is the requirement under section 72 that all carriers be 75 percent Canadian owned. This restriction could be relaxed by Order in Council. This would increase the number of potential entrants and may also increase the discipline imposed on incumbents by the threat of entry. U.S. carriers may be well-placed to enter some domestic Canadian routes or to link various Canadian cities via their U.S. hubs or regional hubs. In addition, U.S. carriers are already realizing frequent flyer, CRS and other network economies. They would have less, if any, need for regulatory intervention to ensure their credibility as entrants.

RESIDUAL RATE OR FARE REGULATION

The National Transportation Agency has the power to control economy fares on southern routes where there is an absence of effective competition. Thus the option of imposing fare ceilings on monopoly routes still exists. The appropriate definitions of a monopoly route and of what constitutes a competitive fare are dealt with below. In this regard the U.S. experience with defining monopoly rail routes and with setting rates on them in the post-*Staggers Act* period is instructive as is the Canadian experience with competitive line rates on railways and with fare regulation on monopoly bus routes.

The *Staggers Rail Act*

The *Staggers Rail Act* left the Interstate Commerce Commission (ICC) with limited jurisdiction over railway freight rates. The ICC can order that a rate be raised to average variable cost which is the minimum "reasonable" rate under the Act. The ICC also has the power to set maximum rates under a restricted set of circumstances. Specifically, the ICC may investigate the reasonableness of a rate if it exceeds 180 percent of variable cost and the carrier has market dominance. If the rate is found to be unreasonable the ICC may order it to be reduced. Any rate below 180 percent of variable cost or which is set by a railway which is not in a position of market dominance is presumed reasonable.

The existence of any one of four forms of competition is sufficient to disprove market dominance. These are:

- intra-modal competition: competition from another railway;
- intermodal competition: competition from road or water transport;

- geographic competition: the ability of a receiver to obtain the product to which the rate in question applies from another source or of the shipper to ship it to another destination; and
- product competition: the ability of the receiver to use a substitute for the product to which the rate in question applies or the ability of the shipper to make and ship a substitute.

Railways deemed to be “revenue inadequate” may further increase their rates by 4 percent annually even if these rates exceed 180 percent of variable cost, and market dominance exists. This is called the “zone of rate flexibility.” At present all U.S. railways are deemed by the ICC to be revenue inadequate, that is, earning less than their cost of capital.

In its assessment of the reasonableness of rates which exceed the 180 percent of variable cost, for which there is market dominance and which are not in the zone of flexibility, the ICC uses a “constrained market pricing” approach. Constrained market pricing is based on two principles. These principles are “differential pricing” and “stand-alone cost.” The principle of differential pricing is that a greater proportion of common costs should be recovered from the customers with the less elastic demands. Differential pricing is thus a simplified version of Ramsey pricing. Because prices exceed marginal cost by the largest amount in the market segments in which this has the smallest effect on demand, Ramsey pricing minimizes the distortion in output associated with the recovery of a given level of fixed costs.

The stand-alone cost principle is that no shipper should pay a rate higher than the lowest stand-alone average cost of serving that shipper or any group of shippers. This is what any shipper would pay if the market for freight services were contestable. That is, given contestability, a railway that charged shippers more than the stand-alone average cost of serving them would lose the market to a new entrant.

The application of these two principles was illustrated by Willig and Baumol.¹⁰⁶ Suppose that a railway serves markets S and T. Variable costs are \$80 in each market, and there are common costs of \$40. Suppose further that as a result of intermodal competition customers in the S market can only be charged \$80. Then customers in the T market pay \$120 which is the variable cost of service plus the fixed costs. This \$120 is also the cost a new entrant would incur in serving the T market alone.

Roberts pointed out that the practical application of the stand-alone average cost ceiling to captive railway shippers presents a number of difficulties. First, it may be difficult to determine what the costs of a specific component of an integrated rail network are. Second, it may also be difficult to determine the group of shippers a hypothetical new entrant might serve.¹⁰⁷

There has been considerable dissatisfaction among captive U.S. shippers with the ICC's constrained market pricing approach. Mid-western coal shippers have been the most vocal. Dunbar and Mehring found, however, that at least during the early years of the *Staggers Act*, real, distance-corrected coal rail rates fell on average.¹⁰⁸

Roberts has proposed an alternative to the stand-alone cost approach to determining rate ceilings.¹⁰⁹ He suggests that all shippers paying rates above the system average price–marginal cost ratio be treated as a class. The rates paid by this group would all have the same ratio to marginal cost. This ratio would be sufficient to cover all fixed system costs. In the terms of Willig and Baumol, the hypothetical entrant is the same as the railway's existing network and the hypothetical group of customers is all customers currently discriminated against (in the sense of paying above-average price-cost margins).

Competitive Line Rates on Canadian Railways

Sections 134–143 of the *National Transportation Act, 1987* provide for the establishment by the National Transportation Agency of competitive line rates. A competitive line rate may be established between either the point of origin or the point of destination of a shipment and a connecting rail carrier. A competitive line rate cannot be applied over more than 50 percent of the distance between the points of origin and destination or 750 miles whichever is greater. The formula to be applied in determining the competitive line rate is given in section 137. The competitive line rate is the inter-switching rate plus the revenue per mile of the connecting carrier multiplied by the number of miles over which the competitive line rate is to apply less the distance over which the interswitching rate applies.

The first competitive line rate case decided by the National Transportation Agency under the 1987 Act is described by Lande.¹¹⁰ The case involved shipments of methanol by tank car from Medicine Hat to Shelby, Montana.

The shipper, Alberta Gas & Chemicals Ltd. was captive to CP Rail at the point of origin but could interswitch with Burlington Northern. Alberta Gas had a rate from Burlington Northern but could not agree with CP on a rate on its portion of the route. The Agency set the CP rate at the Burlington Northern rate per ton-mile in Canadian dollars times the CP mileage in excess of the interswitching distance plus the interswitching charge. This rate has been renewed twice by the Agency.

Extra-provincial Bus Transport

Section 195(2) of the *National Transportation Act, 1987* states that the National Transportation Agency "may, by order, disallow any tariff of rates or any portion thereof if, in its opinion,

- (a) the tariff or portion thereof is not compensatory and is not justified by the public interest; or
- (b) there is no effective alternative and competitive transportation service by a common carrier . . . and the tariff of rates unduly takes advantage of a monopoly situation."

A 1986 Canadian Transport Commission decision under section 40 of the 1967 Act illustrates the factors likely to be considered in determining whether a bus fare is unduly monopolistic. In *Re TerraTransport* (Decision number MV-40-224), the Motor Vehicle Transport Committee of the Commission found no justification for disallowing a 5 percent increase in fares on the Roadcruiser bus service in Newfoundland.¹¹¹ Its reasons were that TerraTransport was not and had not been profitable, there was no evidence of excessive or misapportioned costs or of an overall deterioration in service, and there was no indication that past fare increases had themselves discouraged traffic appreciably. The Committee also implied that it would continue to look favourably on fare increases at TerraTransport until the latter was earning "a reasonable rate of return which would generate enough capital to reinvest in the operation in order to maintain and, if there is sufficient demand, to expand the service."¹¹²

9. IMPLICATIONS OF NORTHERN ROUTES AND OTHER MODE REGULATION FOR OVERSIGHT OF FARES ON MONOPOLY AIRLINE ROUTES IN CANADA

DEFINING A MONOPOLY AIRLINE ROUTE

The factors which might be considered in defining the routes on which the carrier(s) might have market power include:

- the number of carriers on the route (intra-modal competition);
- the ability of additional carriers to offer service on the route (barriers to entry or contestability);
- the number of carriers serving the same city-pair by different routes (interhub competition); and
- the existence and time and cost comparability of alternative modes of transportation including automobile, bus and train (intermodal competition).

There are a number of alternative standards for regulatory intervention. If the objective is to confine regulatory intervention to cases in which the exercise of monopoly power is highly probable, the city-pairs selected for scrutiny (either reactive or proactive) would be served by a single carrier, have significant barriers to entry, and have ineffective interroute (including interhub) or intermodal competition.

Factual evidence of barriers to entry would include restrictions on the access of potential entrants to physical and intangible infrastructure, specifically:

- lack of access to counter or gate space or arrival/departure slots;
- lack of access to interlining privileges;
- lack of (or inferior) access to computer reservation services; and
- lack of access to frequent flyer programs.

Factual evidence would also include the characteristics of potential entrants. Morrison and Winston define a potential entrant as a carrier serving both the origin and destination points on a route but not the route itself.¹¹³ It would be of considerable interest to know how many single carrier routes

in Canada have such “first tier” potential entrants. It should also be noted in passing that the statistical results of Morrison and Winston imply that, other things being equal, the fare on a route declines as the number of “first tier” potential entrants increases but that this relationship is not statistically significant.¹¹⁴ This might be taken to imply that the constraint imposed on incumbents by “second tier” potential entrants who serve either the point of origin or the point of destination of a route but not both and not the route itself is minimal.

More theoretical types of evidence would involve an examination of the characteristics of the route involved and how these might affect the reaction of an incumbent carrier to entry. Do sunk costs and/or regulatory restrictions on exit commit incumbents to the route? Incumbents may be particularly difficult to dislodge from routes where either the origin or the destination is their hub.¹¹⁵ Are fixed costs such that the incumbent(s) and the entrant cannot both earn normal profits on the route? If so, excess profits need not induce entry even if access to the requisite infrastructure does exist.

Evidence regarding the competitiveness of alternative modes might include their respective frequencies, connections, fares and travel times.

DEFINING A REASONABLE FARE

Having established the type of route on which fare regulation has a reasonable probability of being beneficial, the task is then to establish the criteria for adjudicating fares. One approach would be to set fare caps. These caps could take the form of those presently available to captive shippers under the *Staggers Act*. Fares on monopoly routes would be subject to maximum price–variable cost ratios. The maximum ratio would be such that passengers on the route in question pay no more than the stand-alone average cost of their service.

The approach taken by the National Transportation Agency to the reasonableness of northern air fares is somewhat different. The Agency calculates service-specific and possibly route-specific rates of return as well as system rates of return. These calculations assume that both overhead costs and capital can be apportioned to a particular service or route. This implies that overhead services and the services of capital equipment can be rented in small amounts (that is, an airplane for two days per week or ground

personnel or equipment for two hours per day). In contrast the stand-alone cost approach allows for indivisibilities in various functions. In other words, a service may have to bear the entire cost of an airplane if the latter cannot be rented elsewhere when it is not in use on that service.

The Agency's northern fare investigations have not needed to address the issue of differential pricing. Would the Agency allow an excess return on a particular route if the system as a whole was earning normal or less than normal returns? This is likely to be the issue confronting fare regulation on southern monopoly routes in the next few years.

The stand-alone cost approach taken by the ICC appears preferable here. The National Transportation Agency's practice of allocating overheads on a *pro rata* basis implies equal price-variable cost margins across routes or services. While this may have political appeal, it is not what would happen in a contestable market with indivisibilities. If the regulatory process is to approximate the competitive and allocatively efficient outcome, fares should be allowed to exceed variable cost by the amount of fixed unit stand-alone cost. This is likely to differ considerably from route to route.

The stand-alone cost approach rules out the approach to rate-making embodied in the competitive line rate. The imposition of a competitive line rate would imply a requirement that the fare per mile on the monopoly portion of a flight be the same as on any connecting competitive portions. Given differences in equipment and traffic densities there is a strong likelihood that a fare specified in this manner would not cover stand-alone average cost.

It should be noted that, given the vast and variable array of fares now offered on most routes, the imposition of a regulatory ceiling on the economy or Y-fare protects a relatively small fraction of air travellers. Monopoly power can be exploited by manipulation of the discount structure and by withdrawal of discount fares which "forces" travellers to fly at the Y-fare or not at all. The importance of discriminatory pricing poses a considerable regulatory problem. Regulation of a Y-fare at which few people fly is somewhat redundant. Regulation of discount structures and their availability is likely to be complex and costly for all concerned. The controversy over the regulation of freight rates on bulk coal which ought to be a relatively simple task should serve as a warning to those contemplating the regulation of the full spectrum of fares and their availability on monopoly routes.

Moreover, discrimination can result in the marginal user being served at marginal cost and is thus allocatively equivalent to competition under some circumstances. The efficiency consequences of regulating the entire spectrum of fares on monopoly routes are ambiguous at best and are, in all probability, negative once the cost of the regulatory apparatus is taken into account. This leaves Y-fare regulation as a kind of “life line” regulation which guarantees a “reasonable” full cost fare which is always available although it may be infrequently used.

ACCESS REQUIREMENTS

In its provisions relating to rail regulation, the *National Transportation Act, 1987* gives the National Transportation Agency the authority to require that other railway companies be accorded various forms of access to shippers who would otherwise be captive to a single railway. Some of these access provisions may be applicable to air transportation. They might then be employed to increase the degree to which monopoly routes are contestable. These access provisions and Canadian experience with them are described below.

The Act contains a number of provisions which allow a shipper served by a single carrier to have access to competing carriers. These access provisions are as follows:

- *Running rights:* Under section 148 a railway company may, with the approval of the National Transportation Agency, use the whole or any portion of the tracks, terminals, stations or station grounds of any other railway company and exercise full rights and powers to run and operate its trains over any portion of the railway of another railway company. The Agency may fix, by order, the privileges and obligations of each railway as well as the required amount of compensation.
- *Joint use of right-of-way:* Under section 149, the Governor in Council may order the joint or common use of the right-of-way of a railway by two or more railway companies if it is satisfied that this would result in significant efficiencies and reductions in cost. The Governor in Council may also fix by order the extent of this joint or common use and the compensation that must be paid.

- *Connection of intersecting railway lines:* Under section 150, the National Transportation Agency may order that railway lines that intersect, cross or run through the same urban or industrial area be connected to permit the safe and convenient transfer of traffic between the lines and the inter-switching of traffic between the railways involved. The Agency may determine and apportion the cost of effecting and maintaining this connection. Under section 151, the Agency may, in conjunction with the relevant provincial regulatory board, also order the connection of intersecting, provincially regulated railways.
- *Interswitching:* Under section 151(1), where a line of one railway company connects with the line of another railway company, the Agency may, on the application of the company owning either of the railways involved or of a municipal corporation or other interested party, order the companies that operate those lines to afford all reasonable and proper facilities for the safe and convenient interswitching of traffic between those lines. Under section 151(2), where the point of origin or destination of the traffic is within 30 kilometres of the interconnection point (the interswitching limits) the terms and conditions under which interconnection occurs must be as determined by the Agency. The Agency may also extend the interswitching limits if the point of origin or destination is “reasonably close” to the point of interchange.

With respect to the application of the access provisions of the Act, the Agency has made one order under section 150 requiring that two railway lines be interconnected. In 1990 it ordered that the Port Stanley Terminal Railway be connected with CP trackage at St. Thomas, Ontario.¹¹⁶

A number of requests for running rights under section 148 have been submitted to the Agency. One involved an application by CP Rail for running rights over 10 miles of CN trackage in order to serve two large chemical producers in the Fort Saskatchewan Alberta area. The two railways ultimately settled the matter privately.¹¹⁷ Another application filed under section 134 of the 1967 Act involved the use of facilities, namely, the CP passenger station in Regina by VIA Rail. This was also settled privately. Three applications for running rights are currently under review by the Agency. These involve two short-line railways in Ontario and a proposal to run highway trailers on bogies on CN tracks between Moncton and Windsor.

Interswitching has been a feature of Canadian railway regulation since 1908. The National Transportation Agency sets interswitching rates and adjudicates applications for extension of the interswitching limits (currently within 30 kilometres of the interchange point) within which regulated terms and conditions apply. The Agency recently denied an application by CIL of West Carseland, Alberta to be included within the interswitching radius of Calgary. Its reason for denial was that CIL was not disadvantaged with respect to its freight costs.

Regulation may enhance the bargaining power of passengers on monopoly air routes in the same way it has enhanced the bargaining power of captive rail shippers. The analogy between the rail access provisions in the *National Transportation Act, 1987* and possible access provisions for city pair airline markets is somewhat strained but it does serve to illustrate the alternatives.

Running Rights or Joint Use of Right-of-Way

The city-pair airline market analogy to these rail access provisions is the right to use the incumbent's ticketing, baggage handling or servicing facilities and personnel at either or both the point of origin or destination. The moral hazard problems could be such as to require a great deal of regulatory supervision. The problem would be less severe if these services were provided by outside contractors. Access could then be facilitated by prohibiting exclusive contracts. Flint notes that one feature of airline deregulation in Australia has been to require the existing carriers (Ansett and Australian) to make gates that would otherwise be tied up in long-term leases available "at market lease rates" to other carriers.¹¹⁸ Flint goes on to note, however, that this involves only two gates each at Sydney and Melbourne and one gate each at Adelaide, Perth and Coolangatta, and this is unlikely to facilitate entry on a scale sufficient to discipline the incumbents.

An extreme example of running rights would be the right to sell seats on the incumbent's aircraft. The analogy here would be swaps in the petroleum refining industry. Carriers could swap blocks of seats on various routes and price their respective blocks as they see fit. As Williamson suggested in the context of the petroleum refining industry, the swap would also attenuate the incentive to treat the other carrier's passengers poorly.¹¹⁹ Indeed, to the extent that the various ground services can be swapped, the regulatory problem of ensuring an adequate standard of performance may be considerably reduced.

Interconnection and Interswitching

The airline analogy to interconnection (and this is extreme) is to require an incumbent to provide convenient connecting flights at either the origin or destination point to be served by a new entrant. A less extreme analogy would be to prohibit an incumbent from rescheduling connecting flights solely to discourage a new entrant. Another possibility would be to oblige incumbents to allow new entrants access to their frequent flyer programs.

The analogy to interswitching would be to interline baggage, to offer joint rates for ongoing passengers and to hold flights as they would for their own interconnections where flights are intended to connect.

Airlines and railways differ in the extent of the control they exert over their respective infrastructures. Railways own virtually all their rights-of-way, trackage, terminals and other facilities. Regulatory intervention is occasionally required to ensure their availability. Airlines in Canada do not own terminal facilities. These facilities are largely government owned. Access to them is a matter of operating policy rather than regulatory intervention. The possible contribution of terminal operating policy to the contestability of individual airline routes is considered below.

Decentralized Public and Private Responses

Public policies that might increase the contestability of monopoly air routes include:

- ensuring that arrival/departure slots, gates and ancillary facilities at government-operated terminals are available to potential entrants;
- auctioning off the rights to routes that are contestable *ex ante* but weakly contestable or noncontestable *ex post*; and
- recognizing the complementarity of airline and terminal services and other local activities, and providing incentives to reflect it in air fares.

With respect to slot availability, either time-of-day pricing or a system of tradeable slots would put entrants and incumbents on the same footing.¹²⁰ This would be true even if incumbents were grandfathered. There might still be a problem of strategic refusals to sell adjacent departure space to close rivals. This could be dealt with in several ways. The first is to reduce the

strategic and, indeed, the pecuniary value of slots by expanding airport capacities. The second is to treat the monopolization of slots by a carrier or group of carriers acting jointly the same as the monopolization of any other good or service would be treated, that is, to apply the abuse of dominant position or conspiracy sections of the *Competition Act* to it. Similar considerations apply to the provision of gate space and ancillary airport facilities.

Routes which can support only one carrier and are weakly contestable *ex post* may be more strongly contestable *ex ante*. It may be possible to auction off the rights to single carrier routes. The *National Transportation Act, 1987* presently provides for the auctioning of loss-making routes under section 85(2). In this case bids compete to provide a service with the lowest possible subsidy. In the case of single-carrier profit-making routes, bids could take the form of lump-sum fees, economy fares, fare formulas or some combination thereof. Lump-sum fees could be rebated to the communities involved, perhaps credited against municipal taxes.

Route auctions would face two types of problems. The first is that property rights in routes do not exist. Having sold the right to serve it, a community could not prevent the entry of a competing carrier. One possible means by which exclusivity might be assured is by municipal control over the community airport. In this case a community could theoretically guarantee exclusivity to the winning bidder by denying landing rights to potential competitors. This type of a guarantee may not be enforceable and may not provide much comfort to the winning bidder.

The second type of problem is that auctioning off the right to serve is itself a costly process. The resources required to ensure that the winning bidder meets the commitments and to alter the contract as circumstances change can be considerable. Indeed, some authors have argued that when companies are locked in to their arrangement, the process bears a striking resemblance to conventional price, entry and exit regulation.¹²¹

The recognition of the existence of complementarities to air travel and the incorporation of them in fare decisions (that is, their internalization) can result in lower air fares. A simple example is the complementarity between ground services such as parking and terminal concessions, and air services. The sharing of ground service revenue with a carrier reduces its profit-maximizing monopoly fare. This principle can be extended to cover broader

complementarities including the complementarity between air fares and local land values. If an air service confers value on a local industrial park or a convention centre the carrier has an incentive to set fares to encourage traffic, provided it shares in the resulting increase in value of these local assets.

Among the private-sector responses to a situation of non-contestable monopoly would be the exercise of countervailing buyer market power in various forms. One form would be corporate aviation either by charter or by corporate aircraft. A profile of the Canadian corporate fleet and carriers specializing in business charters is presented by Wallace.¹²² Not surprisingly, firms in the resource sector are well represented among both aircraft owners and users of charter services. These same firms may also make extensive use of confidential contracts with scheduled carriers. At present we have no information on the importance or incidence of confidential contracts of this nature. Non-business passengers may be able to exert the same form of leverage by forming charter groups.

Among the private responses to the information problems facing travellers is the investment by carriers, tour booking companies and travel agents in their reputation. Similarly the carriers protect themselves against possible exploitation by sharing ownership of CRSs. Consumers can also take prudent actions with respect to some concerns. For example, bankruptcy is more of a threat to a traveller under the current system. A traveller can pay for a ticket with a credit card. Some issuing companies will suspend the charges while they, rather than the consumer, deal with the bankrupt carrier. Trip cancellation insurance may also help. On the other side of the transaction, many airlines in financial difficulties are forced to take precautionary steps to assuage the concerns of travellers. For example, Eastern in the United States created a \$50 million escrow fund to guarantee tickets in the period before it stopped flying.

10. CONCLUSIONS

The domestic airline industry has changed profoundly since deregulation. There is a consensus that this change has been for the better. The power of entry, exit and pricing freedom to weed out inefficient practices has been demonstrated. At the same time the expectation of economists that city-pair airline markets would be both unconcentrated and easily entered has been

disappointed. Market power is likely to exist on a number of southern Canadian routes. Deregulation constitutes a substantial improvement over the previous regime, but the close approximation to a competitive market that many of its advocates envisaged has not materialized. As a consequence, the full potential benefits of free competition have not been realized.

The fundamental policy question is whether there is a set of policies which would increase the economic benefits derived by Canadians from airline deregulation. There is a clear role for merger policy here. Specification of the precise ways in which merger or other competition policies should be applied is beyond the mandate of this study. A few general comments, however, are in order.

First, the *Competition Act* is well-suited to deal with airline mergers. Evidence of the relationship between the number of competitors and price and quality performance, while not without ambiguities and methodological problems, is both more plentiful and of higher quality than in most other markets. There is also considerable evidence regarding the source and magnitude of operating economies at the route and network levels. The analysis of the trade-off between the effects of a substantial lessening of competition and the efficiency gains resulting from an airline merger required under sections 93–96 of the *Competition Act* is likely to be more complete than in most other markets.

Second, the remedies available under the *Competition Act* could be augmented by further deregulation. The achievement of further network economies while preserving competition would be facilitated if foreign carriers could be licensed to operate on domestic routes. As it stands, there is a danger that the only available remedy will be a quasi-regulatory fix which the National Transportation Agency is arguably better qualified to impose and enforce than the Competition Tribunal. An example of a quasi-regulatory remedy is the Competition Tribunal's consent order in the *Reservac* case.¹²³

There may also be a role for access policies. The most important of these are inherent in the operating procedures of the major airports. Arrival/departure slots and ancillary facilities should be priced at cost and available to incumbents and entrants on similar terms. Access to the network facilities of incumbent airlines could be enforced by the Agency. The access remedies available to railways under sections 148–152 of the *National Transportation*

Act, 1987 may also be available to airlines under sections 77–79 of the *Competition Act*. These are the abuse of dominant position sections of the Act. The withholding of access to various network facilities to single-route entrants may be regarded as an abuse of dominant position. Similarly the Competition Tribunal may find that incumbents have an obligation to supply network facilities to entrants under section 75 of the *Competition Act*. The Agency is probably better equipped than the Tribunal to take on this task, especially if it involves ongoing supervision. An alternative is to allow for competition by other networks. This would, again, require that U.S. carriers be granted domestic licences. The resulting inter-network competition is preferable, in our view, to imposing access requirements either by means of competition or regulatory policy on a monopoly Canadian-owned network. The possibility of network competition and the formation of alliances also reduce any barrier to entry resulting from frequent flyer plans.

There is little to be said for residual fare regulation, especially under current circumstances. Regulation of the basic, or Y-fare, on monopoly routes may meet a political need. It is unlikely to have much impact on the average cost of travel on these routes. Regulation of the full slate of fare offerings is likely to be a very costly undertaking.

Finally, there may be merit in structuring the regulatory framework so that decentralized arrangements can evolve. This may involve giving individual communities the right to negotiate with potential carriers and award exclusive rights to the carrier promising the best service package. It could involve contractual arrangements which leave travel agents with the incentive to choose the lowest fare for their clients. It may also include technical innovations which make computer reservation systems more readily accessible. Regulators must ensure that they do not preclude the emergence of superior decentralized solutions to problems posed by market power.

ENDNOTES

1. Although Canadian regulation was invasive in all these dimensions, the degree of control varied and differed from that exercised in the United States at that time:

A crucial feature of Canada's regulatory approach was its non-directiveness. While fare increases were disallowed from time to time as unjustified by airline costs and financial returns, and while deep discount fares were sometimes disallowed for carrying too few travel and purchase restrictions, the airlines were never told what fares they should charge instead.

In this respect, the Canadian approach was always very different from the U.S. approach. As a result, airline pricing in Canada was determined to a greater extent than in the U.S. by airline managements, rather than by regulators. Unlike its U.S. counterpart before deregulation, the Canadian price control mechanism had no built-in incentives to the airlines to provide excess capacity. (Hans Lovink, "When to Deregulate the Airline Industry: Canada's Approach" in Bureau of Transport Economics, *Economic Regulation of Aviation in Australia, Seminar Papers and Proceedings* (Australian Government Publishing Service, 1985), p. 96.)
2. The Committee endorsed the thrust of the position paper, offering suggestions for details in implementation in its Sixth Report, entitled *Freedom to Move: Change, Choice, Challenge*, Ottawa, 1985.
3. Richard M. Soberman, "Canadian Passenger Transportation Policy," in *Canadian Transportation Policy*, edited by David W. Gillen (Kingston: Queen's University, John Deutsch Institute, 1990), p. 8.
4. Starting at the west coast, the dividing line follows the 55th parallel until Manitoba which it crosses at a diagonal until it reaches the 50th parallel. From this point east the boundary follows the 50th parallel.
5. According to section 67 of the Act this percentage can be reduced by the Governor in Council.
6. CP Air began with "Courieur" fares and followed with "Skybus" fares. AC countered with "Nighthawk" and the "One Way/One Day Skysave" programs.
7. Earl Thompson in "The Perfectly Competitive Production of Collective Goods," *Review of Economics and Statistics* 50 (1968), pp. 1-12 and Harold Demsetz in "The Private Production of Public Goods," *Journal of Law and Economics* 13 (1970), pp. 293-306 reached different conclusions on this matter. The two views are reconciled in Thomas Borcherding, "Competition, Exclusion, and the Optimal Supply of Goods," *Journal of Law and Economics* 21 (1978), pp. 111-32. The Thompson view depends on discrimination being maintained in the face of competition. This debate is related to that which distinguishes between the nature of competition among rivals providing the goods. The outcome will have desirable properties if the competition to sign up the customers is Bertrand competition. If the competition is Cournot, competitive entry will be excessive. See Daniel F. Spulber, *Regulation and Markets* (Cambridge, Mass.: The MIT Press, 1989), p. 42 and references cited there. (See section 6 of this report for an explanation of "Bertrand" and "Cournot" competition.)
8. See T. E. Duffy and P. R. Berlinguette, *The Low-Priced Air Fare Review: The First Five Years* (Canadian Transport Commission, Research Branch, 1983/05e, November 1983).

9. Alfred E. Kahn, "Applications of Economics to an Imperfect World," *American Economic Review* 69 (May 1979), pp. 11–12.
10. Alfred E. Kahn, "Surprises of Airline Deregulation," *American Economic Review* 78, no. 2 (May 1988), p. 320.
11. E. E. Bailey, and J. R. Williams, "Sources of Economic Rent in the Deregulated Airline Industry," *Journal of Law and Economics* 31 (1988), p. 188. © 1988 by The University of Chicago. All rights reserved. 0022-2186/88/3101-0006 \$01.50.
12. As described in Geoffrey Rowan, "Airlines try to lure bargain travellers," *The Globe and Mail* [Toronto] May 17, 1991, pp. B1 and B2.
13. Charters still play a role, although a diminishing one, in the Toronto–Vancouver run. See the National Transportation Agency, *Annual Review 1989* (Ottawa: Supply and Services Canada, 1990) p. 42.
14. "How Airlines Deregulation Can Mean Dollar Savings for Business Travelers," *Dun's Business Month*, April 1985, p. 81.
15. *The Globe and Mail*, [Toronto], February 12, 1991, p. C1.
16. The National Transportation Agency *Annual Review 1989*, p. 43 reported: "On 32 routes which Air Canada and Canadian turned over to their regional affiliate airlines, the number of discount fares more than doubled in 1988 and remained at this level in 1989, averaging over eight per route."

With reference to the availability of discounts on short-haul versus long-haul flights:

In 1986, discount carriage on long-haul services in the southern domestic market represented 66.7% of passenger volume and 67.6% of passenger-kilometres; this compares with 50.7% and 52.5% respectively for short-haul. This distinction was even more pronounced in the preceding three years. Statistics Canada, *Touriscope, 1988 Tourism in Canada: A statistical digest* (Ottawa: Supply and Services Canada, Catalogue No. 87-401, October 1988), p. 110.

17. Statistics Canada, *Air carrier operations in Canada: October–December 1989* (Supply and Services Canada, Catalogue No. 51-002, vol. 20, no. 4, December 1990), Figure 1, p. xi.
18. Samuel Skinner, U.S. Secretary of Transportation, *Airline Marketing Practices: Travel Agencies, Frequent-Flyer Programs, and Computer Reservation Systems*. A study of the Task Force on Competition in the U.S. Domestic Airline Industry, (Washington, D.C.: U.S. Department of Transportation, 1990), p. 31.
19. Information from an interview (October 19, 1990) with Roger Roy of the National Transportation Agency of Canada.
20. Samuel Skinner, *Airline Marketing Practices*, pp. 8 and 9.
21. "AT&T has stated that its auditing service saved the corporation \$700,000 a year even though the company's agency failed to book the lowest fare only two percent of the time." *Ibid.*, p. 18.
22. Guerin-Calvert notes: "CAB found that the U.S. CRSs systematically used their market power in CRS to limit the ability of new entrants (carriers either new to the industry or to dominated markets) to get their lower fares to the public, that new carriers' costs were

- increased, and that customers paid substantially higher prices for airline service than they would have paid in the absence of CRS market power." See Margaret E. Guerin-Calvert, "Competitive Analysis of the Reservec-Pegasus Merger," March 2, 1989, Affidavit submitted by Director of Investigation and Research, Bureau of Competition Policy to the Competition Tribunal, Ottawa, in the *Gemini* case, p. 37.
23. Kenneth W. Thornicroft, "Airline Deregulation and the Airline Market," *Journal of Labor Research*, 10, no. 2 (Spring 1989), p. 166.
 24. Thomas G. Moore, "U.S. Airline Deregulation: Its Effects on Passengers, Capital, and Labor," *Journal of Law and Economics*, 29 (1986), p. 26.
 25. Bailey and Williams, "Sources of Economic Rent," p. 193.
 26. Jordan has argued that, despite the agitation in 1979 and 1980, a longer perspective indicates that strikes/lockouts played a smaller role in the U.S. in the first six years of deregulation than they did in the last four years of regulation. William A. Jordan, "Results of U.S. Airline Deregulation: Evidence from the Regulated Canadian Airlines," *Logistics and Transportation Review* 22, no. 4 (1986), p. 325.
 27. *The Vancouver Sun*, December 1, 1989, pp. C1 and C5.
 28. *The Globe and Mail*, [Toronto] October 10, 1990, pp. B1 and B2.
 29. *Ibid.*, October 13, 1990, pp. B1 and B2.
 30. Jean-François Bence, "La déréglementation du transport aérien," *Policy Options Politiques* 11, no. 2 (March 1990), p. 15.
 31. Kahn, "Surprises of Airline Deregulation," p. 319.
 32. National Transportation Agency of Canada, *Annual Review 1989*, Figure 3.9, p. 42.
 33. John R. Meyer and Clinton V. Oster Jr., *Deregulation and the Future of Intercity Passenger Travel* (Cambridge, Mass.: The MIT Press 1987), p. 124.
 34. Moore, "U.S. Airline Deregulation", p. 21.
 35. Steven A. Morrison and Clifford Winston, "Enhancing the Performance of the Deregulated Air Transportation System," in *Brookings Papers on Economic Activity: Microeconomics*, edited by M.N. Baily and C. Winston (Washington, D.C.: The Brookings Institution, 1989), p. 73.
 36. The potential of hub-and-spoke architecture was realized before deregulation but its introduction required the freedom to realign routes which deregulation provided. A reference for an engineering description of the advantages of hubbing is Adib Kanafani, "Aircraft Technology and Network Structure in Short-Haul Air Transportation," *Transportation Research* 305, 15A, (1981). An early discussion of the economics appears in Arthur S. De Vany and Eleanor H. Garges, "A Forecast of Air Travel and Airport Use in 1980," *Transportation Research* 6, 1 (1971). In conversation with the authors, George Hariton, who was formerly the Director of Research at the Canadian Transport Commission, noted that Boeing was an important force in developing the hub-and-spoke system.

37. *Flight International*, June 3, 1989, pp. 29–32.
38. To gain perspective on Toronto's North American status as an airport, it is worth noting that in 1988 it was the 21st busiest airport in the world on the basis of passengers embarked and the 20th busiest on the basis of aircraft movements. International Civil Aviation Organization, *Civil Aviation Statistics of the World 1988*, DOC 9180/14 (1989), pp. 150–51.
39. Statistics Canada, *Touriscope, 1988 Tourism in Canada: A statistical digest*, p. 107.
40. As quoted by Lovink, pp. 99–100.
41. David W. Gillen, T. H. Oum and M. W. Tretheway, "Airline Cost Structure and Policy Implications," *Journal of Transport Economics and Policy* (January 1990), p. 28. What is considered a cost factor or a demand factor depends on the definition of output. Gillen et al. felt that such factors as frequency of flights and more attractive frequent flyer programs, affect demand. We view these as elements of quality which the smaller producer finds more costly to provide.
42. Douglas W. Caves, Laurits R. Christensen and Michael W. Tretheway, "Economies of Density Versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ," *Rand Journal of Economics* 15, no. 4 (Winter 1984).
43. Jordan, "Results of U.S. Airline Deregulation," p. 301.
44. David R. Graham and Daniel P. Kaplan, *Competition and the Airlines: An Evaluation of Deregulation*. Staff Report of the Office of Economic Analysis, Civil Aeronautics Board (Washington, D.C.: December 1982), p. 46.
45. Gillen, Oum and Tretheway, "Airline Cost Structure," n. 15.
46. In the United States, the exception is Southwest Airlines, which has been successful and is expanding. However, it has stuck to a regional strategy, adopted point-to-point scheduling rather than a hub configuration, and has not chosen to compete directly with the major carriers on the main routes between large cities.
47. Kahn, "Surprises of Airline Deregulation," p. 318. Kahn is perhaps unique in organizing the demise of the regulatory body, the Civil Aviation Board, which he headed.
48. Before the strike there were 17,725 controllers of which 13,300 were full-time. In March 1987 there were 14,900 controllers of which 11,200 were full-time.
49. The controlled airports currently are Chicago O'Hare, Washington National and Kennedy and LaGuardia in New York.
50. "During the worst two-day period in December, 111 flights into or out of Pearson were cancelled, causing disruption of air traffic from coast-to-coast." National Transportation Agency of Canada, *Annual Review 1988*, p. 50.
51. Garth Wallace, "Living with Slots," *Aviation and Aerospace* (January 1990), p. 51.
52. Robert W. Hahn and Randall S. Kroszner, "The Mismanagement of Air Transport: A Supply-Side Analysis," *The Public Interest* 95 (Spring 1989), pp. 105–106.

53. Slots not used 65 percent of the time in a two-month period are withdrawn by the FAA; international slots are not covered by the rule.
54. See Table B-4 of Samuel Skinner, U.S. Secretary of Transportation, *Airports, Air Traffic Control, and Related Concerns (Impact on Entry)*. A study of the Task Force on Competition in the U.S. Domestic Airline Industry, (Washington, D.C.: U.S. Department of Transportation, 1990).
55. *Ibid.*, p. 2-9.
56. *Ibid.*, p. 2-17.
57. *Ibid.*, p. 2-15.
58. National Transportation Agency of Canada, *Annual Review 1989*, p. 27.
59. Morrison and Winston, "Enhancing the Performance," p. 93.
60. *The Globe and Mail*, [Toronto], November 20, 1990.
61. National Transportation Agency of Canada, *Annual Review 1989*, pp. 38, 40.
62. Moore, "U.S. Airline Deregulation," p. 27.
63. T. H. Oum, W. T. Stanbury and M. W. Tretheway, "Airline Deregulation in Canada and its Economic Effects," Working Paper #90-TRA-013, Draft, June 29, 1990, p. 19.
64. *The Globe and Mail*, [Toronto], October 12, 1990, pp. A1 and A2.
65. National Transportation Agency of Canada, *Annual Review 1989*, p. 30.
66. "Service levels in Canada's top 20 city-pairs increased substantially during the year, as evidenced by a 14 per cent increase in the number of direct flights and a 24 per cent increase in the seat capacity offered (fourth quarter 1988 versus fourth quarter 1987)." National Transportation Agency of Canada, *Annual Review 1988* (Ottawa: Supply and Services Canada, 1989) p. 57.
67. Oum, Stanbury and Tretheway, "Airline Cost Structure," Exhibit 5.
68. Federal Trade Commission, *The Deregulated Airline Industry: A Review of the Evidence* (1988) n. 3, p. 10.
69. National Transportation Agency of Canada, *Annual Review 1989*, p. 30
70. *Ibid.*, p. 55.
71. *Ibid.*, pp. 29-30.
72. Bence, "La déréglementation," p. 16.
73. An s-curve describes the relationship between carrier capacity and revenue on a route. Presumably, more frequent flights increase the value of a ticket on any one flight since they give more options should customers have to make adjustments in their travel times and because of frequent flyer programs.

74. From an Exhibit in the USAir–Piedmont Merger Case, July 1987 as cited in Severin Borenstein, "On the Efficiency of Competitive Markets for Operating Licences," *Quarterly Journal of Economics* 103 (1988), p. 357.
75. Kahn, "Surprises of Airline Deregulation," p. 318.
76. Lawrence F. Cunningham, M. B. Slovin, W. R. Wood and Janis K. Zaima, "Systematic Risk in the Deregulated Airline Industry," *Journal of Transport Economics and Policy*, (September 1988), p. 347.
77. Thomas Moore reported in "U.S. Airline Deregulation: Its Effects on Passengers, Capital, and Labor," "In spite of the bankruptcy of Braniff and the poor economic performance of Continental, TWA, and Western, the real value of the ten trunk airlines, including Pan Am, has remained virtually unchanged from December 1, 1976, to December 1, 1983. For comparison, real value of all stocks on the New York Stock Exchange fell 3 percent over the same period. Some airlines, such as American, which tripled in value, did very well in this new unregulated market, as measured by the stock value of the company. But the big gainers were the regional carriers. The thirteen regional carriers witnessed the value of their companies balloon almost sixfold in real terms."
78. See "Airlines fly towards financial abyss," *The Ottawa Citizen*, June 19, 1990, p. D7.
79. Fred Lazar, in his *Deregulation of the Canadian Airline Industry: a Charade* (Toronto: Key Porter Books, 1984), p. 46, stated:

Pricing wars and destructive competition are to be expected in an industry in which all firms produce identical products, there is no dominant price leader, fixed costs comprise the largest share of total costs and demand declines.
80. Lee J. Van Scyoc, "Effects of Airline Deregulation on Profitability," *The Logistics and Transportation Review* 25, no. 1 (March, 1989), p. 48.
81. The margin is the difference between operating revenues and operating expenses (excludes depreciation and interest charges).
82. Level I airlines are the larger carriers. From 1981 to 1986 the class included AC, CP, EPA, Nordair, Quebecair, PWA and Wardair. In 1988, AC, CAIL and Wardair represented the class. Currently there are two Level I carriers, AC and CAIL.
83. See G. J. Stigler, "A Theory of Oligopoly," *Journal of Political Economy* 72, no. 1 (February 1964), pp. 44–61.
84. The model employed here is an address model in which carriers have different locations in "departure space" but are otherwise identical. For a general discussion of address models see Steven Salop, "Monopolistic Competition with Outside Goods," *Bell Journal of Economics* 10, (1979), pp. 141–56.
85. See for example National Transportation Agency of Canada, Staff Report, File Number DR-88 3004, Appendix III, (1988), p. 5.
86. Joseph A. Schumpeter, *Capitalism, Socialism and Democracy* (New York: Harper & Row, Torchbook edition, 1962), p. 85.
87. With increasing average costs, the incumbent could not prevent a subset of customers from signing up with an entrant, and the break-even price is not sustainable. However,

- within the capacity of an airplane, average costs per passenger decline, and the carrier's break-even price would be sustainable.
88. Technically, sustainability in this context means that the charges to each group must be in the core of the price game. It is conceivable that the core is empty, but the more likely outcome is that more than one assignment will be in the core. Under quite general circumstances Ramsey prices (prices which maximize welfare subject to a break-even constraint) are in the core.
 89. Soberman, "Canadian Passenger Transportation Policy," p. 5.
 90. Robert W. Hahn and Randall S. Kroszner, "The Mismanagement of Air Transport: A Supply-Side Analysis," *The Public Interest* 95 (Spring 1989), p. 105.
 91. Lazar, *Deregulation*, p. 137.
 92. Michael W. Tretheway, "Analysis of the Effect of the Gemini Computer Reservation System Merger on Competition in the Canadian Airline Industry," (March 1, 1989). Affidavit submitted by Director of Investigation and Research, Bureau of Competition Policy in the *Gemini* case, p. 19. Note that CAIL has only purchased a minority interest in its feeders.
 93. *Ibid.*
 94. The average round trip was 1,179 miles. Morrison and Winston, p. 67 and n. 15.
 95. David W. Gillen, W. T. Stanbury and Michael W. Tretheway, "Duopoly in Canada's Airline Industry: Consequences and Policy Issues," *Canadian Public Policy* 14(1) (March 1988), p. 27.
 96. Morrison and Winston, "Enhancing the Performance," p. 83 and fn. 44.
 97. *Ibid.*, p. 67.
 98. The European and American codes cover display bias, pricing, contract length, terms and conditions of contracts, and provide non-discriminatory access to enhancements such as direct access links. Margaret E. Guerin-Calvert, "Competitive Analysis," p. 46.
 99. Michael E. Levine, "Evidence," (March 1, 1989), submitted by Director of Investigation and Research, Bureau of Competition Policy to the Competition Tribunal, Ottawa, in the *Gemini* case, p. 16.
 100. Skinner, *Airline Marketing Practices*, p. 6.
 101. Bailey and Williams, "Sources of Economic Rent," p. 189.
 102. Skinner, *Airline Marketing Practices*, p. 15.
 103. See *The Globe and Mail* [Toronto], December 15, 1989, p. B6.
 104. Asra Q. Nomani. "Fare Game: Airlines may be Using a Price-Data Network to Lessen Competition," *Wall Street Journal* [New York], June 28, 1990, p. A1.
 105. Kahn, "Surprises of Airline Deregulation," p. 319.

106. R. D. Willig and W. J. Baumol, "Using Competition as a Guide," *Regulation* (1987), no. 1, pp. 28-35.
107. M. J. Roberts, "Residual Rate Control: The Unmet Challenge of Deregulation," *The Logistics and Transportation Review* 23, 1 (1987) pp. 83-108.
108. F. Dunbar, and J. Mehring, "Coal Rail Prices During Deregulation: A Hedonic Price Analysis," *The Logistics and Transportation Review* 23, 1 (1987) pp. 17-34.
109. Roberts, "Residual Rate Control," n. 107.
110. R. Lande, *Railway Law and the National Transportation Act* (Toronto: Butterworths, 1989), pp. 140-41.
111. Canadian Transport Commission, Decision number MV-40-224 (MV-86-1) (Ottawa, 1986).
112. *Ibid.*, p. 35.
113. Morrison and Winston, "Enhancing the Performance," pp. 61-112.
114. *Ibid.*, p. 74.
115. Morrison and Winston found that, other things being equal, fares were higher on routes where either the origin or the destination was the hub of a major carrier. See Morrison and Winston, *ibid.*, p. 74. Others dispute the proposition that routes involving hubs are particularly difficult to enter. See "The Myth of the Fortress Hub," *The Avmark Aviation Economist* (October 1990), pp. 6-9.
116. National Transportation Agency of Canada, *Annual Report 1990* (Ottawa: Supply and Services Canada, 1991), p. 20.
117. Lande, *Railway Law and the National Transportation Act*, p. 147.
118. P. Flint, "Toto, I have a feeling we're not in Kansas anymore," *Air Transport World* (May 1990), pp. 52-59.
119. O. E. Williamson, "Credible Commitments Using Hostages to Support Exchange," *American Economic Review* 73 (September 1983), pp. 519-40.
120. Borenstein has pointed out that if bidders in a slot market differ in their ability to capture surplus slots, the free market allocation of slots may not maximize total surplus. This need not imply any asymmetry between incumbents and entrants. See S. Borenstein, "On the Efficiency of Competitive Markets for Operating Licenses," *Quarterly Journal of Economics* 103 (May 1988), pp. 357-85.
121. See O. E. Williamson, "Franchise Bidding for natural monopolies — in general and with respect to CATV," *The Bell Journal of Economics* 7 (Spring 1976), pp. 73-104.
122. See G. Wallace, "Flying for Business," *Canadian Aviation* (October 1989), pp. 23-31.
123. The Consent Order specifies detailed rules for the display of flight information on computer reservation systems and for the relationships between these systems and carriers and subscribers respectively. See "In the matter of an application by the Director of Investigation and Research under subsection 64(1) of the *Competition Act*, R.S.C. 1970, c. C-23, as amended; Consent Order (Competition Tribunal, CT - 88 / 1).

CONSTITUTIONAL JURISDICTION OVER TRANSPORTATION: RECENT DEVELOPMENTS AND PROPOSALS FOR CHANGE

Patrick J. Monahan*

December 1991

I. INTRODUCTION

The purpose of this study is twofold: to provide an overview of constitutional jurisdiction in the field of transportation, and to consider the need for constitutional change in this field.

The study examines the relevant provisions in the Canadian Constitution allocating jurisdiction over the field of transportation,¹ and considers the judicial interpretation of these provisions and the extent to which the courts have modified or supplemented the original scheme contemplated by the Constitution. It then examines how legislative jurisdiction has actually been exercised by both federal and provincial governments to establish whether the constitutional division of responsibilities has limited or constrained the ability of governments to respond to changing circumstances in transportation.

This analysis leads logically to the second purpose of the study — a consideration of the need for constitutional change in the field of transportation. Does the existing constitutional scheme require amendment? Such an analysis is timely, given the current constitutional discussions that are ongoing in

* Director, York University Centre for Public Law and Public Policy.

Canada. In September of 1991, the Government of Canada initiated a process aimed at a fundamental and comprehensive re-examination of the Canadian Constitution.² In addition to the proposals put forward by the government, there have been a variety of other suggestions for fundamental constitutional change published over the past year.³ Many have included proposals for change to the constitutional allocation of jurisdiction over transportation. It is important that the debate over such changes be grounded in a detailed and concrete understanding of the existing scheme of the Constitution; it is also important to identify with some care the possible implications of any proposed constitutional amendments. This study will attempt to provide such an understanding.

The first section of the paper outlines the existing constitutional jurisdiction in the field of transportation. Since this is an area that has been discussed quite extensively in the academic literature,⁴ the emphasis here is on recent developments, particularly the recent Supreme Court of Canada judgements which have clarified the respective roles and responsibilities of the federal and provincial governments in relation to transportation. This section outlines the various provisions in the Constitution which grant federal or provincial jurisdiction over transportation and reviews the regulatory framework which both levels of government have put in place on the basis of their constitutional responsibilities. Finally, it identifies the issues or areas which can be expected to generate litigation in the future, as reflected by an analysis of recent court decisions.

The first issue discussed in the second section of this study is whether the existing constitutional framework has led to any obvious problems or difficulties. The study examines the extent to which the current constitutional framework may have prevented governments or the private sector from responding to changing needs or circumstances in transportation. It also analyzes the extent to which the Constitution may impede government as it faces the challenges inherent in designing a transportation system that will meet future needs of Canadians.⁵

The second section of the paper also considers the merits and implications of a variety of proposals to change constitutional responsibilities in relation to transportation. Over the last year, the Constitutional Committee of the Quebec Liberal Party (The Allaire Committee),⁶ the Group of 22,⁷ as well as the Government of Canada,⁸ have all put forward proposals for altering the

Constitution as it relates to transportation. This section also considers the extent to which the existing Constitution permits governments to delegate or alter regulatory responsibility in the transportation field.⁹ The paper concludes with an overall assessment of the need for formal constitutional amendments in the field.

II. THE EXISTING CONSTITUTIONAL FRAMEWORK

The *Constitution Act, 1867* does not classify transportation as a class of subject (or head of power) assigned exclusively to Parliament or the provincial legislatures. Instead, specific transportation matters or modes are dealt with in a variety of separate constitutional provisions which effectively divide responsibility for transportation regulation between the federal and provincial governments. In general terms, the Act allocates jurisdiction over inter-provincial and international transportation to the federal government, while reserving to the provinces responsibility for transportation matters within a single province. This territorial approach to transportation¹⁰ is reflected most clearly in section 92(10) of the *Constitution Act, 1867*, which reserves to the federal Parliament responsibility over “Works and Undertakings connecting the Province with any other or others of the Provinces, or extending beyond the Limits of the Province,”¹¹ while providing for provincial responsibility for “Local Works and Undertakings.”

Other provisions in the 1867 Act which allocate jurisdiction to the federal Parliament include sections 91(9) (“Beacons, Buoys, Lighthouses, and Sable Island”); 91(10) (“Navigation and Shipping”); 91(13) (“Ferries between a Province and any British or Foreign Country or between Two Provinces”); section 92(10)(c) (power to declare local works for the “general Advantage of Canada”); and section 108 (certain Public Works and Property of each Province was transferred to Canada, including Canals, Public Harbours, Railways and Military Roads).¹² The federal power over “trade and commerce” in section 91(2) of the 1867 Act was at least potentially relevant to the field of transportation. The courts have construed this provision narrowly, however, and it has never been interpreted as adding significantly to federal authority in this field.¹³ Federal authority over criminal law in section 91(27) has also permitted the federal government to establish a set of criminal prohibitions and sanctions relating to the operation of motor vehicles, vessels and aircraft.¹⁴ Provincial authority in relation to transportation matters flows

from section 92(10) ("Local Works and Undertakings"); section 92(13) ("Property and Civil Rights in the Province"); and section 92(16) ("Matters of a merely local or private Nature in the Province").

The courts have also been called upon to supplement the original division of powers contemplated by the 1867 Act as new modes or methods of transportation arise. Of greatest significance in this regard is air travel which of course was unknown in 1867 and was therefore not mentioned in the original division of powers. The courts have interpreted the federal Parliament's power to make laws for the "Peace, Order and good Government of Canada" as including the exclusive authority to regulate all aspects of air travel.

The terms of section 92(10), establishing federal jurisdiction over inter-provincial works and undertakings, have been the greatest single source of constitutional litigation in the field of transportation. The principles which the courts have developed in their interpretation of this provision make up the essential core of the constitutional jurisprudence in the transportation field.

JURISDICTION OVER WORKS AND UNDERTAKINGS

Section 92(10) of the *Constitution Act, 1867* provides that the provincial legislatures have exclusive power to make laws in relation to:

Local Works and Undertakings other than such as are of the following Classes: —

- (a) Lines of Steam or Other Ships, Railways, Canals, Telegraphs, and other Works and Undertakings connecting the Province with any other or others of the Provinces; or extending beyond the Limits of the Province:
- (b) Lines of Steam Ships between the Province and any British or Foreign Country:
- (c) Such works as, although wholly situate within the Province, are before or after their Execution declared by the Parliament of Canada to be for the general Advantage of Canada or for the Advantage of Two or more of the Provinces.

Although section 92(10) is in terms a grant of legislative power to the provinces, the exceptions established in subsections (a), (b) and (c) have proven to be the most significant feature of the provision. These exceptions from provincial authority represent grants of exclusive legislative authority to the Parliament of Canada, in accordance with section 91(29) of the *Constitution Act, 1867*.

There are a number of settled principles with regard to the interpretation of section 92(10). The first relates to the distinction between “Works” and “Undertakings” referred to in the provision. The courts have interpreted an “Undertaking” as involving both a physical and an organizational element. Viscount Dunedin in the *Radio Reference* referred to an undertaking as “not a physical thing, but . . . an arrangement under which . . . physical things are used.”¹⁵ Thus, in the *Winner* case, the federal Parliament possessed jurisdiction not only over the buses which provided the interprovincial transportation, but also over the bus company itself.¹⁶ This functional approach means that federal authority over interprovincial undertakings extends to all aspects of the organization or enterprise which provides the service in question.

A second settled principle has to do with the fact that constitutional jurisdiction over a particular work or undertaking is to be undivided: for the purposes of section 92(10) jurisdiction is allocated to a single level of government. The courts have consistently rejected the idea of dividing jurisdiction between the federal and the provincial governments over a single undertaking. This fundamental principle was first established in the *Bell Telephone* case of 1905.¹⁷ The Privy Council rejected the idea that the telephone company’s long-distance business and its local business should be separated for the purpose of allocating legislative jurisdiction. The Board held that the telephone company was engaged in an interprovincial undertaking and thus the whole of the company’s business, including its strictly local activity, fell under federal jurisdiction.

This approach is quite different from that adopted by the Privy Council in relation to its interpretation of the federal trade and commerce power in section 91(2) of the *Constitution Act, 1867*. It consistently restricted the federal authority over trade and commerce to the interprovincial or international aspects of trade; the local aspects of trade remained subject to exclusive provincial jurisdiction and could not be reached by federal legislation. Thus,

in the *Natural Products Marketing Reference* (1937),¹⁸ a federal statute regulating natural products that were primarily traded in international markets was ruled invalid since the statute included some transactions which could be completed within a single province. The reasoning of the Privy Council was that federal authority could only be exercised in relation to those transactions which crossed provincial borders. This segmented approach to the construction of the trade and commerce power has proven to be one of the key factors in limiting the scope and usefulness of this particular source of federal authority.¹⁹

On the other hand, the Privy Council's determination that jurisdiction over transportation undertakings was to be undivided has led to quite different results in the transportation field. Once an undertaking is classified as interprovincial, federal jurisdiction immediately extends to all aspects of the enterprise, including any features that are strictly local. This has meant that federal authority to regulate transportation undertakings has been much more extensive and therefore more effective than in many other areas of federal jurisdiction. In particular, the Privy Council's undivided approach to transportation undertakings has meant that this is one of the few areas in which the federal government is capable of effective action without the necessity of involving provincial governments.²⁰

The courts' resistance to dual jurisdiction in the transportation field has had important implications in the central issues which have emerged in litigation surrounding section 92(10). The allocation of jurisdiction has been treated by the courts as "an all or nothing affair";²¹ a transportation undertaking is subject *either* to federal jurisdiction or to provincial jurisdiction, but not simultaneously to both. This has meant that the key question for purposes of section 92(10) has been the characterization of an undertaking as either local or interprovincial. This, in turn, has led to two recurring questions that continue to dominate the court decisions in this area:

1. What is the extent of the interprovincial activity or connection that is necessary to support a finding that a given undertaking is interprovincial or international as opposed to local?
2. To what extent can federal jurisdiction be extended to an otherwise purely local undertaking because that local undertaking is functionally integrated or connected with an interprovincial undertaking?

INTERPROVINCIAL WORKS AND UNDERTAKINGS

The courts have established a relatively low threshold of interprovincial activity to support a finding that a particular undertaking qualifies as an interprovincial one. The courts have consistently held that an undertaking falls within federal regulatory authority even if only a small percentage of its business activity is interprovincial or international. The primary test is whether the interprovincial or international services are a “continuous and regular” part of the undertaking’s operations. If this requirement is met, then the whole undertaking is subject to exclusive federal regulation.

There are many examples of this rule being applied so as to include primarily local undertakings within federal jurisdiction. In the case of *Re Tank Truck Transportation* (1960),²² the issue was whether the *Ontario Labour Relations Act* was applicable to an Ontario trucking company whose operations were predominantly confined to the province of Ontario. The evidence before the Court was that, in 1959, the trucking firm had completed 94 percent of its trips within the province, with just 6 percent extending beyond provincial borders. But the Court found that the interprovincial activity was a “continuous and regular” aspect of the trucking firm’s operations and, as a result, the whole of the undertaking, including the local operations within Ontario, was subject to the exclusive authority of the Parliament of Canada.²³ A similar ruling was made in the *Liquid Cargo* case (1965),²⁴ where only 1.6 percent of a trucking firm’s trips extended beyond provincial boundaries.

The most recent Supreme Court of Canada pronouncement on this issue was made in *Alberta Government Telephones v. Canadian Radio-television and Telecommunications Commission and CNCP Telecommunications*.²⁵ The question here was whether Alberta Government Telephones (AGT), a provincial Crown corporation operating a telephone system in Alberta, fell under federal or provincial authority. AGT’s physical facilities were located entirely within the province of Alberta and the system could carry telephone messages only within the province. However, the AGT system was connected with other telephone companies outside the province to enable local subscribers to make extra-provincial telephone connections. AGT argued that it fell under provincial regulatory authority since its activities were confined totally to the territory of the province of Alberta. The Supreme Court unanimously rejected this claim, holding that AGT was subject to exclusive

federal authority.²⁶ In reaching this conclusion, the Court articulated a number of general principles that it indicated ought to guide analysis of this issue:

1. The location of the physical apparatus of an undertaking in a single province and the fact that all the recipients of a service are within a single province will not preclude a finding that an undertaking is interprovincial in scope. The primary concern is "not the physical structures or their geographical location, but rather the service which is provided by the undertaking through the use of its physical equipment."²⁷
2. In considering the nature of the service or operation, one must look to the "normal or habitual activities of the business as those of 'a going concern', without regard for exceptional or casual factors. . . ."²⁸
3. It is impossible to formulate in the abstract a single comprehensive test which will be useful in all cases; instead, the court must be guided by the "particular facts in each situation. . . ."²⁹

Applying these principles to the situation of AGT, the Court found that the operations of the Crown corporation were interprovincial and international in scope. The primary basis for this conclusion seemed to be that AGT provided a service which enabled residents of Alberta to communicate beyond the borders of the province. According to Chief Justice Dickson, who wrote for the Court on this point, "AGT is, through various commercial arrangements of a bilateral and multilateral nature, organized in a manner which enables it to play a crucial role in the national telecommunications system."³⁰ It was the capacity to provide this extra-provincial service which supported the finding that AGT was subject to exclusive federal authority.

One leading commentator, Professor P. Hogg, has observed that this represents a more expansive reading of federal authority than has been adopted in other contexts.³¹ As he has pointed out, the fact that a local undertaking is capable of providing a service beyond the borders of a single province had previously been regarded as an insufficient basis for asserting federal regulatory jurisdiction. For example, in the *Cannet Freight Cartage* case,³² a freight forwarder provided local customers with the opportunity to ship goods beyond the borders of the province. The freight forwarder took delivery of goods in one province and made all the arrangements necessary to ship the goods to another province by rail. The Ontario Court of Appeal

found the freight forwarder to be subject to exclusive provincial jurisdiction because its own operations were limited to a single province. The Court reasoned that the freight forwarder did not become an interprovincial undertaking by virtue of shipping goods on an interprovincial railway.

Professor Hogg has expressed the view that it is not easy to see a difference in the facts that make up the *AGT* case and those of earlier cases such as *Cannet Freight*. He has suggested that what might explain the Supreme Court's most recent decision was the sheer scope and complexity of the agreements between AGT and the other Canadian telephone companies. These multilateral agreements meant that AGT was part of what amounted to an integrated national telecommunications network.³³ What seems evident, in any event, is that the Supreme Court was prepared to take a slightly broader view of federal regulatory authority in this case than it had previously. It is also significant that the Court was prepared to move in the direction of greater federal authority in an area which had traditionally been subject to control by the provinces.

Historically, federal regulatory authority had included Bell Canada (serving Ontario and Quebec), the British Columbia Telephone Company, as well as telephone companies serving Yukon, the Northwest Territories and parts of Newfoundland.³⁴ But the telephone companies in the other provinces had traditionally been subject to provincial or local control.³⁵ Thus, in a practical sense, the Court's decision in *AGT* had quite significant practical implications. It opened the door for federal regulatory authority in a context which had traditionally been regarded as subject to exclusive control of the provinces.³⁶

INTEGRATION OF LOCAL AND INTERPROVINCIAL UNDERTAKINGS

As noted, a transportation undertaking can be classified as federal if the undertaking itself is regarded as interprovincial (as in the *AGT* case), or if a purely local undertaking is integrated or connected with another undertaking which is itself interprovincial.

The precise degree of the connection or integration that is required has been the subject of extensive litigation over the years. An early Privy Council case determined that mere physical connection between a local railway and an interprovincial railway was insufficient to bring the local railway under federal authority.³⁷ Integration in an operational or functional sense is

required before local undertakings fall within federal authority. For example, a local railway line that was operated under a formal management agreement by the CNR was held to fall within federal authority.³⁸ Similarly, a company supplying stevedore services in Toronto to seven shipping companies involved in international shipping was held to be subject to federal labour legislation.³⁹ Although the stevedore company was independent of the shipping companies, the services which it provided were integral to the successful operation of the shipping enterprises. The same reasoning was applied in the *Letter Carriers* case,⁴⁰ in which a trucking company which had contracted with the Post Office to deliver and collect mail was found to be within federal authority. The Court found that the trucking operation was integral and necessary to the operations of the Post Office itself.

The most recent Supreme Court of Canada judgement on this issue is the *Central Western Railway* case.⁴¹ The issue here was whether Central Western Railway, a small railway located entirely within the province of Alberta, fell within federal or provincial jurisdiction. Central Western used its 105 miles of track in central Alberta to transport grain from nine elevators to the CNR's interprovincial rail line. The grain cars were then transported by CN to Vancouver for export. Central Western's tracks were separated from those of CN by a four-inch gap, and the CNR controlled the device which regulated entry onto its line. The issue for the Court was whether the degree of connection and integration between Central Western and CNR was sufficient to subject the local railway to federal jurisdiction.

Chief Justice Dickson, speaking for a majority of the Court,⁴² rejected the argument that Central Western could be regarded itself as an interprovincial railway, noting that mere physical connection between a local and an interprovincial rail line was an insufficient basis for establishing federal jurisdiction. He cited the *AGT* case, arguing that "[t]he linchpin in the *AGT v. C.R.T.C.* decision was this court's finding that AGT, by virtue of its role in Telecom Canada and its bilateral contracts with other telephone companies, was able to provide its clients with an interprovincial and, indeed, international telecommunications service."⁴³ The Chief Justice regarded Central Western's operation as quite different from that of AGT. He noted that Central Western simply moved grain within Alberta and that the interprovincial transportation of grain was handled entirely by CN. On this basis, he concluded that Central Western was a local railway and not itself part of an interprovincial undertaking.

The Chief Justice then turned to the second possible basis for finding in favour of federal authority. Even though Central Western was a local railway, it would fall under federal authority if it could be characterized as an integral part of a federal work or undertaking. Dickson C. J. indicated that this integration might develop in at least two different ways.⁴⁴ First, the management and operation of Central Western might be coordinated or undertaken in common with that of an interprovincial undertaking. Second, the effective operation of a federal undertaking might be dependent on the services of Central Western.

Dickson C. J. concluded that Central Western was not functionally integrated with any interprovincial undertaking and therefore not subject to federal authority. He reasoned that Central Western and CN were operated as separate undertakings rather than in common; further, CN was not dependent on the services of Central Western for its own operations. Nor was Central Western integrated in a functional sense within a so-called "Western Grain Transportation Network."⁴⁵

Despite the Court's ruling in this particular case, Chief Justice Dickson's judgement illustrates the very broad reach of federal regulatory authority in this field. The functional character of the Court's approach is noteworthy. Even where there is in form two separate undertakings, the courts will inquire into the degree of practical or operational integration between the undertakings. Federal regulatory authority will extend to any operations that are regarded as essential or conducted in common with a core interprovincial undertaking. The practical effect of this approach is to ensure an expansive interpretation to federal authority under sections 92(10)(a) and (b) of the *Constitution Act, 1867*.

WORKS DECLARED FOR THE GENERAL ADVANTAGE OF CANADA

Paragraph 92(10)(c) of the 1867 Act provides an exception to the principle that local works and undertakings are subject to exclusive provincial jurisdiction. This subsection provides that the Parliament of Canada may simply declare that a local work is "for the general advantage of Canada or for the advantage of two or more of the provinces"; such a declaration is sufficient to bring an otherwise local work within federal regulatory authority.

This power has been used close to 500 times, mostly in the late 19th and early 20th centuries and in most cases in relation to railways.⁴⁶ It can be used in relation to a specific work or to a broad class of works.⁴⁷ Moreover, the “works” in question need not be limited to the field of transportation or communication but can involve any sort of physical or tangible thing.⁴⁸ Once the declaration is issued, the courts will not inquire into whether the work is in fact for the general advantage of Canada. The declaration by Parliament will be regarded by the courts as dispositive.⁴⁹

Various commentators have suggested that the declaratory power is inconsistent with classical principles of federalism since it permits the federal government to increase its jurisdiction unilaterally at the expense of the provinces.⁵⁰ The power has fallen into relative disuse and appears to have been used only twice in the last 25 years.⁵¹ A recent federal proposal that the power be abolished⁵² is discussed later in this study.

PEACE, ORDER AND GOOD GOVERNMENT

The opening words of section 91 of the *Constitution Act, 1867* grant the Parliament of Canada power “. . . to make laws for the Peace, Order and good Government of Canada, in relation to all Matters not coming within the Classes of Subjects by this Act assigned exclusively to the Legislatures of the Provinces.” While this source of federal authority has generally been interpreted narrowly by the courts, one important exception has been in the field of transportation. The “Peace, Order and good Government” power has been held to support exclusive federal jurisdiction over air transportation.⁵³ The Supreme Court of Canada found aeronautics to be a “distinct subject matter” which went beyond local concern and “must from its inherent nature be the concern of the Dominion as a whole.”⁵⁴

The effect of the *Johannesson* case was to include all aspects of aeronautics, including purely local aeronautics undertakings, as subject to exclusive federal authority. Thus the distinction between interprovincial and local undertakings, which has been critical in the judicial interpretation of section 92(10), has no application to the field of aeronautics. Even purely local airline operations fall under exclusive federal regulatory authority, without any requirement that the local operation be connected to or integrated with an interprovincial undertaking.⁵⁵

Given the settled nature of federal jurisdiction over aeronautics, the litigation in this field has tended to focus on a variety of subsidiary issues such as the extent to which provincial laws of general application apply to airports. The courts have tended to hold that airports and aeronautics undertakings are exempt from the application of any provincial legislation which affects a vital part of the federal undertaking. For example, it has been determined that airports are exempt from municipal zoning bylaws of general application⁵⁶ as well as from height restrictions imposed by a province on land adjacent to an airport.⁵⁷

A second issue that has produced some litigation is the extent to which federal jurisdiction extends to undertakings that are connected to aeronautics. Here, the courts have relied on the jurisprudence developed in relation to section 92(10)(a); the issue has been whether the related undertaking is sufficiently integrated with the main aeronautics undertaking. For example, in the *Field Aviation* case,⁵⁸ the Alberta Court of Appeal held that a company engaged in the servicing of aircraft was so intimately connected with aeronautics as to fall within federal jurisdiction. On the other hand, a company constructing airport runways,⁵⁹ as well as companies offering porter services or limousine service to and from the airport⁶⁰ have been held to be separate undertakings subject to provincial jurisdiction.

OTHER CLASSES OF SUBJECTS

Three other enumerated classes of subjects assigned to Parliament in section 91 of the *Constitution Act, 1867* deal explicitly with matters related to the field of transportation.⁶¹

These are:

91.9 Beacons, Buoys, Lighthouses and Sable Island

91.10 Navigation and Shipping

91.13 Ferries between a Province and any British or Foreign Country or between Two Provinces.

Federal authority over transportation matters is also supplemented by section 108 of the *Constitution Act, 1867*, which provides as follows:

108. The Public Works and Property of each Province, enumerated in the Third Schedule to this Act, shall be the Property of Canada.

The Third Schedule includes the following classes of provincial public works and property:

1. Canals, with Lands and Water Power connected therewith.
2. Public Harbours.
3. Lighthouses and Piers, and Sable Island.
4. Steamboats, Dredges, and public Vessels.
5. Rivers and Lake Improvements.
6. Railways and Railway Stocks, Mortgages, and other debts due by Railway Companies.
7. Military Roads.

Although the Third Schedule transfers to the federal government all improvements or public works associated with rivers and waterways, the ownership of the rivers themselves remains with the provinces.⁶² This means that the provinces may legislate with respect to the use of these waters, as long as their legislation does not interfere with federal legislation in relation to navigation and shipping.

The most important source of federal authority from the above catalogue of powers is section 91(10), "Navigation and Shipping". The language in the section is unqualified, suggesting that federal authority could be extended to all aspects of this subject. However, for many years the courts seemed to take the position that this head of power was circumscribed by the same limits that had been developed with respect to federal undertakings under section 92(10)(a).⁶³ In *Agence Maritime v. Canada Labour Relations Board*,⁶⁴ for example, it was held that local shipping was subject to provincial labour relations legislation. Similarly, ferries that operated largely within the waters of British Columbia were held to be within provincial jurisdiction for purposes of labour legislation.⁶⁵ In these cases, the courts seemed to interpret federal authority over navigation and shipping as extending primarily to interprovincial and international undertakings.

The most recent decision of the Supreme Court of Canada on this issue suggests a somewhat more expansive reading of federal authority over navigation and shipping. In *Whitbread v. Walley* (1990),⁶⁶ the issue was whether certain limitations of civil liability contained in the *Canada Shipping Act*, R.S.C. 1970, c.S-9 applied to a pleasure boat operated within provincial waters. The Supreme Court unanimously held that the provisions in the *Canada Shipping Act* applied uniformly to all shipping, including local shipping as well as pleasure boats. Mr. Justice La Forest, writing for the Court, distinguished the federal power over navigation and shipping in section 91(10) from that applicable to works and undertakings in section 92(10)(a). Whereas federal jurisdiction over works and undertakings was limited to interprovincial and international transportation, there was no such limitation with respect to navigation and shipping. La Forest J. stated that Parliament's jurisdiction over maritime law should be viewed as territorially co-extensive with its jurisdiction in respect of navigable waterways.⁶⁷ He rejected the idea that any distinction could be made between local shipping and interprovincial shipping. Instead, he took the view that all navigable waterways within Canada are part of a single navigational network which must be subject to a uniform legal regime. La Forest J. drew an analogy between navigation and shipping and the field of aeronautics which, as noted above, has been regarded as a single subject matter within the exclusive authority of Parliament. This is so, Justice La Forest suggested, because it is functionally impossible to make a distinction between air travel of a local versus an interprovincial nature. The same situation holds true, according to Justice La Forest, with respect to navigation and shipping. This points to the need for a uniform regulatory and legal regime for navigation and shipping and for "a broad reading of the relevant head of federal jurisdiction."⁶⁸

The analogy which Mr. Justice La Forest draws between "navigation and shipping" and "aeronautics" is significant. Certainly the federal power over navigation and shipping has never been regarded as being as extensive as the power over aeronautics.⁶⁹ This result is somewhat ironic, given the fact that the area of navigation and shipping is an enumerated head of federal authority, while aeronautics has simply been added through judicial interpretation of the federal residual power. The result and the reasoning in *Whitbread and Walley* indicate the Supreme Court's willingness to reassess this situation and to consider expanding the limits of federal authority over navigation and shipping.

REGULATORY FRAMEWORK

As to the manner in which governments have actually exercised their constitutional authority, our analysis thus far indicates that the courts have taken a much more expansive approach to federal authority over transportation matters than they have in other fields, such as federal authority to regulate trade and commerce.⁷⁰

To what extent does the federal and provincial legislation enacted in this field reflect the fairly centralized scheme contemplated by the formal Constitution? In general terms, the regulatory framework does recognize a leading role for the federal government over transportation matters. However, in certain instances the federal government has chosen not to exercise the full range of authority which it has been allocated under the Constitution. Of course, this is a political rather than a constitutional stipulation, one which can be reversed by ordinary legislation.

What follows is a brief overview of the regulatory framework that has been put in place in the four principal modes of public passenger transportation: air, water, rail and motor vehicle. The focus of this analysis is on particular modes of transportation or forms of regulation which have been an important source of litigation or court decisions in the past. As such, there is no attempt to provide a comprehensive outline of federal or provincial legislation relating to transportation matters.⁷¹ The main purpose of including this information is to provide a more complete understanding of how governments have actually used their formal powers set out in the Constitution.

Air Transportation

The federal government currently dominates the regulation of all air passenger transportation in Canada. Under the *National Transportation Act, 1987* it has exclusive responsibility for regulating the provision of all air services in Canada; and under the *Aeronautics Act* it regulates the safety and security of passengers, aircraft, airport and aviation facilities. All air carriers in Canada are subject to exclusive federal regulation under these statutes, including carriers engaged in purely local transportation. The provincial role in the transportation field is currently limited to establishing and directly operating certain airports as well as subsidizing some air passenger services.⁷² Almost all of the airports and airstrips owned and operated by provincial governments are located in remote, northern areas of the provinces.⁷³

Three provinces directly subsidize air passenger services; one province (Ontario) has established a provincial Crown corporation to provide air passenger services directly in Northern Ontario.⁷⁴ However, all the provincially-operated airports as well as all air passenger services must be federally licensed and meet all the relevant federal regulatory requirements. In short, the regulatory framework governing aeronautics reflects the centralized interpretation developed by the Supreme Court of Canada in this area.

Marine Transportation

The primary public mode of marine passenger transportation is provided by passenger and automobile ferries. The Parliament of Canada has established safety requirements under the *Canada Shipping Act* which apply to all ferry services, including ferries operating within a province. This statute represents a codification of the “rules of the road” for all navigation and shipping within Canadian navigable waters. The federal government also assumes responsibility for the provision of ferry services between provinces,⁷⁵ as well as for certain ferry services that are intra-provincial in nature.⁷⁶ Many of the provinces also provide local ferry services, the most important of these being provided by British Columbia, Ontario, Newfoundland and Quebec.⁷⁷ However, even these provincially operated services are subject to the safety and operational requirements of the *Canada Shipping Act*, thus ensuring a uniform regulatory framework across the country. The provinces have not imposed any additional ferry safety requirements on services which they operate or subsidize but would be free to do so as long as their regulations did not conflict with the paramount provisions in federal law.

Rail Transportation

The federal Parliament regulates the vast majority of passenger rail services in Canada. The federal government has responsibility for all interprovincial and international railways, as well as for any other railways which have been declared to be for the general advantage of Canada. This includes the vast majority of all passenger rail operations in the country. These rail services are regulated by a variety of federal statutes including the *National Transportation Act, 1987* (NTA) and the *Railway Act* (RA). The NTA provides for an administrative agency, the National Transportation Agency, and grants it certain regulatory powers in relation to federally-regulated passenger rail services. Certain decisions of the Agency can be varied or rescinded by the federal Cabinet.

Provincial regulation of passenger rail services is extremely limited. In British Columbia,⁷⁸ Ontario⁷⁹ and Quebec⁸⁰ there are various provincially operated or subsidized local rail services which are subject to provincial regulation. But these provincial services are confined mainly to commuter rail networks or to remote, northern regions. The vast majority of all passenger rail activity is subject to exclusive federal regulation.

Motor Vehicle Transportation

The primary public mode of motor vehicle transportation is provided by the bus industry. As noted above, the Privy Council decision in the *Winner* (*supra* note 16) case established that bus undertakings engaged in regular interprovincial service fell under exclusive federal authority. However, the bus industry had traditionally been regulated at the provincial level, and the federal government had no regulatory structure in place to assume control over the industry. Accordingly, within months of the Privy Council decision in *Winner*, the federal government delegated the regulation of interprovincial undertakings back to the provinces.

The *Motor Vehicle Transport Act, 1987* transfers regulatory authority over interprovincial motor vehicle undertakings to provincially appointed boards. The provincial boards are granted the authority to license the undertakings and to determine the terms and conditions under which they will operate.

It is an established principle of Canadian constitutional law that one level of government cannot directly delegate legislative powers to another level of government.⁸¹ Thus it was inevitable that questions would be raised regarding the validity of the delegation to provincial boards under the *Motor Vehicle Transport Act*. However, the Supreme Court of Canada upheld the validity of the delegation in *Coughlin v. Ontario Highway Transport Board*,⁸² ruling that a provincial board, validly constituted under provincial law to regulate local undertakings, could be vested with the authority to regulate extra-provincial undertakings. This left the regulation of the bus industry to the provinces and has led to significant variations in the applicable regimes governing bus operations across the country.⁸³

Two general points should be noted about the delegation of authority to regulate the bus industry. First, the delegation under the *Motor Vehicle Transport Act, 1987 (MVTA)* can be revoked, altered and limited by ordinary federal legislation. Indeed, there is already a provision in the MVTA

permitting the federal Cabinet to exempt certain undertakings from the Act through regulation.⁸⁴ This exemption has only been used on one occasion, with respect to the Roadcruiser bus service in Newfoundland.⁸⁵ The point is that it can be used at any time in the future by the federal government without the necessity of obtaining provincial consent. This would mean that the federal government could resume responsibility for the regulation of any or all aspects of the interprovincial bus industry simply by passing an Order in Council or by amending the terms of the *Motor Vehicle Transport Act, 1987*.

The second general point is that there appear to be some limits to the capacity of the federal government to delegate authority to the provinces. In *Coughlin* the provincial board that was granted the authority was already validly established under provincial law. Implicit in this decision, therefore, is the requirement that the provincial legislation establishing the provincial board be valid independently of any federal law.⁸⁶ This would mean that the province could not establish a board or agency whose sole purpose was to regulate interprovincial undertakings.⁸⁷ The provincial legislation establishing such an agency would be beyond the constitutional capacity of the provinces, since there would be no valid provincial purpose which the agency was fulfilling.

FUTURE LITIGATION

This discussion indicates that the main outlines of jurisdiction in the field of transportation are now relatively settled. Over the years, the courts have identified a series of general principles which govern the allocation of jurisdiction over transportation. These principles are reasonably well understood by both government and industry, and the dividing line between provincial and federal responsibility is relatively clear.

However, transportation issues will continue to be a frequent source of constitutional litigation in the future. This litigation will arise out of the attempt to apply the general principles identified by the courts to the facts of specific cases. While such issues will vary, one or two central issues will tend to recur with particular frequency.

The first of these issues is the extent to which federal regulation can be extended to local undertakings which have been integrated or connected with interprovincial undertakings. Future transportation systems will place

an emphasis on developing more integrated forms of travel, permitting Canadians to use different modes of transportation on a single trip.⁸⁸ As local and interprovincial modes of transportation become more integrated, the constitutional jurisdiction over these systems might shift in favour of the federal government. As we have seen, the courts have adopted a functional approach to the issue of jurisdiction, inquiring into the degree to which there is common management or operation of various transportation undertakings. As intermodal transportation increases, the issue will be the extent to which an otherwise purely local transportation undertaking becomes integrated within some larger interprovincial network. In such cases, the precise dividing line between federal and provincial responsibility may be uncertain and require clarification by the courts.

A second, related issue which may produce future litigation relates to a possible move away from regulation based on distinct modes of transportation. In the past, governments have tended to establish separate schemes of regulation for specific modes of transportation. In recent years, however, there has been a recognition that certain issues, such as safety, security, substance use and environmental pollution cut across all modes.⁸⁹ The federal government has responded by enacting legislation based on a particular issue or public purpose rather than on a specific mode of transportation.⁹⁰ Within this context, the distinction between “local” and “interprovincial” undertakings is largely meaningless. For example, legislation designed to regulate environmental harm will presumably be aimed at all undertakings with possible adverse effects on the environment. Yet purely local undertakings may cause just as much harm to the environment as interprovincial ones. Thus for such legislation to be truly effective, some means must be found to ensure that all undertakings which produce adverse environmental effects are subject to a set of common standards.

Since constitutional jurisprudence developed by the courts has been premised on a traditional, modal theory of government regulation, the principles which apply to one mode of transportation are quite different from those which apply to others. If the distinctions between modes begin to break down, it may cause the courts to re-evaluate some of its previous jurisprudence. In particular, it may prompt the courts to try to create some room for the development of a set of uniform principles — issue-specific

rather than mode-specific — which would apply across a series of transportation modes. The regulation of adverse environmental effects is a prime instance of where such a re-evaluation might take place.

Even were this to occur, it would be an incremental process which would take many years to develop. The main outlines of constitutional authority over transportation policy are now so well settled, there should not be any significant shift in the courts' approach to these matters in the foreseeable future. The litigation which does arise will be issue-specific and narrowly focussed, and will turn on how the generally accepted principles apply to the facts of specific cases. These fact-specific judicial decisions are unlikely to involve significant implications for transportation policy makers and planners.

III. PROPOSALS FOR CONSTITUTIONAL CHANGE

The first section of this paper has provided a snapshot of the current constitutional arrangements governing the field of transportation. The question which now arises is whether these existing arrangements require modification or amendment.

It is important to keep in mind the distinction between purely statutory arrangements, subject to amendment through ordinary legislation, and the dictates of the Constitution. The focus here is on the degree to which the terms of the formal Constitution require modification. The related issue of whether government policy in the transportation field ought to be altered in some way is beyond the scope of this paper.

This section begins with an examination of the degree to which the terms of the Constitution may have limited governments or the private sector from responding effectively to transportation issues. It measures the practical impact and significance of the Constitution on both government and industry. A number of issues arise here. First, has the Constitution prevented governments from putting in place forms of regulation which would be better suited to the underlying transportation marketplace? In effect, has the Constitution forced governments to rely upon inefficient or ineffective forms of regulation? Secondly, has the Constitution had any negative impact on the

way in which the transportation industry delivers services to the public? For example, has it acted as a barrier to the development of more efficient modes of transportation?

Then follows an examination of various proposals for constitutional change; this assessment focusses on recent proposals for constitutional amendment which have surfaced on the public policy agenda and suggests which, if any, ought to be included in any amendments to the Constitution.

THE PRACTICAL IMPACT OF EXISTING CONSTITUTIONAL ARRANGEMENTS

The transportation industry in Canada has undergone a fundamental transformation since Canada's original constitutional arrangements were put in place nearly 125 years ago. Despite this transformation, the Constitution has not prevented governments from responding effectively to the changing circumstances because of the courts' flexible and functional interpretation of the division of powers in the field of transportation. As the transportation industry has evolved, the courts have in effect updated the Constitution. This has ensured that any gaps or ambiguities are resolved in a satisfactory fashion. It has also meant that constitutional jurisdiction has evolved to take account of changing needs and circumstances in the industry.

The most obvious example of this constitutional "updating" is in air transportation. The courts have allocated exclusive jurisdiction over air transportation to the federal government, thus ensuring the development of an effective national air transport system. But the courts have taken a similarly broad interpretation to federal authority over interprovincial works and undertakings and have found that even predominantly local undertakings are subject to federal regulatory authority if they are engaged in regular interprovincial activity. This preference for federal authority is important, since it permits the creation of a national "level playing field" to govern the transportation industry.

This is not to say, however, that either the federal or provincial governments have put in place a regulatory structure which actually creates this level playing field. The Interim Report of the Royal Commission on National Passenger Transportation makes the observation that the laws affecting passenger transportation are "fragmented" and suggests that passenger transportation companies "may not be competing against other modes on

a level playing field.”⁹¹ However, this fragmentation is more a result of policy decisions by successive governments than by the dictates of the Constitution.

Certainly with respect to the main public modes of passenger transportation — air, bus, marine and rail — the federal government possesses very broad regulatory powers under the Constitution. But the Constitution is largely permissive; it creates room for federal regulatory authority but does not require that authority to be exercised in any particular manner. Thus the federal government has chosen in some instances to delegate at least some of its authority back to the provinces.⁹² In other cases, federal statutes have been drafted in such a way so as not to take complete advantage of the potential federal authority available in a particular area. For example, the *Canadian Transportation Accident Investigation and Safety Board Act* establishes a federal board with authority to investigate and report upon accidents in certain specified modes of transportation.⁹³ The Act does not apply to accidents occurring on highways, where the vast majority of all transportation accidents occur. However, there is certainly no constitutional reason which would prevent the federal government from extending the Act to govern at least some aspects of highway transportation.

As noted above, the federal government has authority over interprovincial undertakings operating on highways. Furthermore, the Parliament of Canada has enacted legislation governing the safety of new cars and components sold in Canada.⁹⁴ In addition, the federal Parliament regulates the behaviour of individual drivers through a variety of provisions in the *Criminal Code* of Canada. These various sources of federal authority would provide an ample basis for extending the application of the federal safety legislation to at least some aspects of highway transportation. Thus the fact that the legislation is framed more narrowly is attributable to factors other than the absence of appropriate constitutional authority in the federal Parliament.

What of the extent to which the Constitution may have indirectly limited the development of more efficient modes of transportation? Again, there seems little basis for concluding that the Constitution has had such an impact. Consider the development of intermodal or integrated systems of transportation, in which passengers might rely on more than one means of travel to arrive at their destinations.⁹⁵ Such systems depend upon the construction

of terminals that serve airplanes, trains, buses and taxis; they also depend upon the development of support services such as intermodal baggage handling, scheduling, reservation systems and ticketing. It might be thought that divided constitutional jurisdiction in the transportation field might make the development of such intermodal systems more difficult to achieve. For example, the fact that the provinces have jurisdiction over local undertakings, while the federal government has jurisdiction over interprovincial undertakings, might be seen as a factor limiting the development of integrated transportation networks.

In fact, however, there would appear to be little basis for supposing that the Constitution would prevent the emergence of more integrated transportation systems. As noted in the first section of this paper, the courts have framed constitutional jurisdiction over transportation undertakings in functional terms. Federal jurisdiction extends to a purely local undertaking only when it is integral to the successful operation of the related interprovincial undertaking or when the two are managed in common. In short, there is no hard and fast distinction between those undertakings that are subject to provincial jurisdiction and those subject to federal authority. The jurisdictional line depends upon the facts and circumstances of particular cases and turns on the degree of functional integration between the various elements of a transportation network.

The greater the degree of functional integration between different modes of transportation, the more likely it will be that the integrated system as a whole will fall under exclusive federal jurisdiction. This is because the integrated or intermodal system will necessarily involve at least some undertakings or modes of transportation that already fall under exclusive federal authority. As these federally regulated modes of transportation become integrated within some larger network of transportation undertakings, the larger system itself may well become subject to exclusive federal jurisdiction. For example, jurisdiction over intermodal stations for buses, trains, urban transit and airplanes appears to be vested in the federal Parliament⁹⁶ because all air travel as well as a significant amount of bus and train travel are already within exclusive federal jurisdiction. The operation of terminals to link these federally regulated modes of transportation with other modes would be an integral part of these federally regulated undertakings. As such, the federal Parliament would possess exclusive authority to regulate the operation of such terminals.

There are, of course, numerous obstacles to the emergence of intermodal transportation systems. The infrastructure to support such systems is extremely costly, while the successful operation of the system can be frustrated by congestion and delays in the urban transportation network.⁹⁷ But these are economic, social and political problems, rather than constitutional ones. An amendment to the Constitution would not make the emergence of intermodal systems significantly more likely or feasible.

Thus, in general terms, I would conclude that the Constitution has not posed a serious problem for either government or industry in the transportation field in the past. This being said, it should also be recognized that there are inevitably going to be certain difficulties associated with divided jurisdiction over transportation matters. The main difficulty has already been alluded to earlier: divided jurisdiction carries with it the possibility that different modes or systems of transportation will find themselves competing on an uneven playing field. If regulations set by one level of government are inconsistent with those set by another, it may produce a situation in which a particular mode of transportation receives a competitive advantage.⁹⁸ This is particularly the case since it is now becoming apparent that competition occurs across transportation modes rather than simply within a particular mode. Within this context, the actions of all three levels of government — municipal, provincial and federal — have impacts on transportation policies undertaken at each level. For example, the choice between air, bus, rail or private automobile for a particular intercity traveller will be affected by such factors as the nature of the road system, the degree of urban congestion or the development of good public transit. Yet these factors are subject to the control of different levels of government. Thus, even though a particular mode of transportation might be subject to the exclusive control of a single level of government — such as aeronautics — the competitive position of that industry will be affected by decisions taken by other levels of government.

Some degree of divided jurisdiction would appear to be inevitable, since it is simply not feasible for a single level of government to assume jurisdiction over all matters which have an impact on the transportation sector. If divided jurisdiction is a fact of constitutional life, there will always be the possibility that the actions of one level of government might frustrate or undermine the policies of another. This problem cannot be solved through a constitutional amendment, since any amendment would still leave a situation in which jurisdiction was divided between two (or perhaps three)

orders of government. The only long-term solution to the problem is to achieve greater coordination and cooperation between the various levels of government.

PROPOSALS FOR CONSTITUTIONAL CHANGE

Over the last year, a variety of proposals for comprehensive constitutional change have been advanced both by governments and non-governmental bodies. While jurisdiction over transportation has certainly not been a central feature of any of these proposals, some of the proposed changes would affect transportation policy.

1. The Group of 22⁹⁹

The Report of the Group of 22 recommends that the declaratory power of the federal government in section 92(10)(c) of the *Constitution Act, 1867* be abolished (recommendation 4). The Report also recommends that inter-provincial and international transportation be federal and intra-provincial be provincial (recommendation 19). Since there is no elaboration of the intent underlying this recommendation, it is not clear whether it contemplates any change to the existing division of powers over transportation. In one sense, the recommendation might be regarded as an affirmation of the *status quo*, since subsections 92(10)(a) and (b) of the *Constitution Act, 1867* provide for federal responsibility for interprovincial undertakings and provincial responsibility for local undertakings. However, it is possible that it contemplates some reallocation of jurisdiction over aeronautics or navigation and shipping, which are now wholly subject to federal authority.

The Report also recommends that the residual power of the federal Parliament be deleted from the Constitution, leaving the question of the allocation of undetermined powers to the political process and the courts. The courts would allocate powers “based on the roles of the two orders of government as reflected in the distribution of powers” (recommendation 3). The effect of this proposal on constitutional jurisdiction over transportation is altogether unclear. Presumably the existing jurisdiction of the federal government over aeronautics, which flows from the federal residual power, would be maintained; it is unclear what would occur with respect to new modes of transportation.

2. The Allaire Report¹⁰⁰

The Constitutional Committee of the Quebec Liberal Party proposes a comprehensive revision of the division of powers. The Committee proposes that transportation be an area of shared jurisdiction between the federal and provincial government, with Quebec having authority over what is termed "regional" transportation, while the federal Parliament would have authority over "inter-regional" transportation (p. 39). There is no indication whether the proposed provincial responsibility over "regional" transportation would involve transportation beyond the borders of the province of Quebec or whether it would be confined purely to intra-provincial transportation. The Allaire Committee also proposes to grant all residual powers to the provinces. There is no indication whether this would affect existing court decisions which have granted jurisdiction to the federal Parliament on the basis of the residual power.

3. Federal Government Proposals

The federal government's proposals for constitutional amendment, entitled *Shaping Canada's Future Together*, contain a number of recommendations with potential impact on transportation. The federal government proposes to abolish the declaratory power with respect to local works in section 92(10)(c) of the *Constitution Act, 1867* (proposal 23). It also proposes to limit the scope of the federal government's residual power (proposal 22). The federal residual power would not be abolished, however; the proposals contemplate that the federal government would retain authority over matters of "national concern," which was the basis for the Supreme Court's decision that aeronautics was a matter of exclusive federal authority. All that is proposed to be transferred to the provinces is authority for "non-national matters not specifically assigned to the federal government under the Constitution or by virtue of court decisions."¹⁰¹ Finally, the federal government identifies a number of areas of jurisdiction which are "candidates for streamlining." The intent of this proposal is to eliminate duplication and overlap of services by the various levels of government. The list of items to be discussed with the provinces includes transportation of dangerous goods, ferry services and small craft harbours.¹⁰²

4. Assessment

Both the federal government and the Group of 22 propose to abolish the federal declaratory power in section 92(10)(c) of the *Constitution Act*,

1867. However, there is no discussion in either document of the rationale underlying this recommendation. One can only assume that it is based on the assumption that a unilateral power of this type is inconsistent with the equality of the two levels of government. Some commentators, as well as provincial governments, have criticized the declaratory power on this basis, arguing that it grants the federal government the power to extend its own jurisdiction unilaterally and reduce the jurisdiction of the provinces.¹⁰³

These criticisms must be balanced against a recognition of the important role which this power has played in the past. For example, shortly after World War II, the federal government used the declaratory power to establish federal jurisdiction over atomic energy.¹⁰⁴ The existence of the declaratory power was important in this context since it permitted swift and effective action on the part of the federal government — a necessary response because of the implications of atomic energy on national security. Speedy and effective national regulation was essential. It is arguable that the courts would have eventually recognized that atomic energy would fall under federal authority as a matter of “national concern.”¹⁰⁵ But even so, there would have been a period of uncertainty during which the status of the federal legislation would have been unclear. The existence of the declaratory power provided a means of virtually eliminating this uncertainty and ensuring effective and timely federal intervention.

Some commentators have suggested that the declaratory power has fallen into disuse and that its abolition would have little practical effect.¹⁰⁶ In fact, while it has been used sparingly in recent years, it was relied on by the federal government as recently as 1987.¹⁰⁷ And it is impossible to predict the kinds of situations or problems which may emerge in the future. The existence of the declaratory power preserves flexibility in the constitutional framework, ensuring the ability of the federal government to respond effectively to changing circumstances.

Despite these considerations in favour of retaining the declaratory power, it would seem preferable to impose some kind of limitation on its use by the federal government. The existence of such a unilateral and unconstrained power is inconsistent with the fundamental equality of the two orders of government.

Furthermore, while the declaratory power has served an important and useful function in the past, it must be remembered that this was during a period in which the Privy Council had adopted a very narrow interpretation

to other sources of federal authority. The federal trade and commerce power, for example, was interpreted by the Privy Council as applying mainly to interprovincial and international trade. In recent years, the Supreme Court of Canada has adopted a much more flexible and expansive approach to the interpretation of the federal trade and commerce power. Thus the need for a unilateral federal power to declare works for the general advantage of Canada has diminished considerably.

It would seem appropriate to limit the use of the declaratory power in order to protect provincial interests. For example, it might be provided that the federal government could only invoke the power after obtaining the consent of the province in which a particular “work” is situated,¹⁰⁸ or that the federal government must obtain the consent of some number of provinces before issuing a declaration.¹⁰⁹ These changes would protect provincial interests while preserving some limited scope for the use of the declaratory power.

Both the Group of 22 and the Allaire Committee propose to eliminate the federal residual power from the Constitution. The federal government, on the other hand, proposes merely to limit the power. The total abolition of the residual power of the federal government might have significant implications in the field of transportation. As noted above, the courts have used the residual power to recognize exclusive federal authority over aeronautics. As new modes or methods of transportation emerge, the courts might well rely on the residual power as a basis for federal authority. To abolish the residual power altogether would appear to represent an unwarranted limitation on the powers of the federal Parliament.

The federal government proposals suggest that the provinces should be granted authority over “non-national matters not specifically assigned to the federal government.” However, the federal government proposes to maintain its authority to deal with “national matters or emergencies.”¹¹⁰ This is important, since federal authority over aeronautics is based on the fact that this is a matter of national concern.¹¹¹ Thus it would appear that the federal proposals on the residual power do not contemplate any changes in the current jurisdiction over aeronautics. More importantly, the federal government could also acquire jurisdiction over modes or methods of transportation which might arise in the future. Jurisdiction over these would turn on the question of whether the transportation matter raised a question of “national concern.”

In short, there does not appear to be any significant difficulty with the current federal proposals in relation to the residual power. The only real area of concern is whether it will be possible to find language that is sufficiently precise to distinguish between matters of “national” versus “non-national” importance. The current proposals require the addition of constitutional language to the opening words of section 91 of the *Constitution Act, 1867* to grant some additional scope for provincial authority. If the language is ambiguous or imprecise, then there could be unintended or unanticipated consequences in terms of the division of powers. Since the federal proposals do not set out the terms of a formal constitutional amendment, it is impossible to state in advance whether this concern is real or merely theoretical.

The other proposals outlined above do not appear to contemplate any significant realignment of responsibilities in transportation. In general terms, they assume the continuation of federal responsibility for interprovincial and international transportation, with the provinces maintaining control over local transportation. However, the federal proposals for streamlining government appear to contemplate some form of delegation of responsibility to the provinces in areas such as ferries and small craft harbours.

In an evaluation of such delegation, it should be noted that the Constitution already permits some degree of delegation of powers between governments. While the courts have held that direct delegations of powers between legislatures are invalid, they have permitted delegations made to third parties, such as administrative tribunals. Thus the federal Parliament has been able to delegate authority to regulate interprovincial motor vehicle undertakings to provincially-appointed regulatory tribunals through ordinary federal legislation — the *Motor Vehicle Transport Act, 1987*.

The immediate question is whether this existing delegation power might be employed to transfer regulatory authority over other transportation matters to provincial authorities. For example, could regulatory authority over interprovincial ferry services be transferred to a provincially appointed tribunal through an ordinary federal statute similar to the *Motor Vehicle Transport Act, 1987*? Alternatively, could jurisdiction over an interprovincial rail undertaking, such as a high-speed train operating in Ontario and Quebec, be transferred to the provinces?

The answer to these questions would depend on the precise nature of the delegation which was contemplated. The delegation in the *Motor Vehicle Transport Act, 1987* involved a transfer of authority to provincial boards which were already validly constituted under provincial legislation. It would appear that the federal government could make a similar delegation involving ferries, railways or any other mode of transportation. The only necessity would be that the provincial boards which were to exercise the delegated power be validly constituted under provincial law. The province or provinces would have first to create tribunal(s) with authority to regulate intra-provincial and local undertakings. These provincial boards could then receive delegated power from the federal government to regulate interprovincial undertakings operating into or within the province concerned.

There are obviously important limitations and drawbacks associated with a delegation of this type. Specifically, any delegation would have to involve several different provincial boards located in different provinces. These separate provincial boards would only be able to regulate the operations of an interprovincial transportation undertaking within the borders of a particular province. To give a concrete example of how this would work, suppose the federal government wanted to delegate authority to regulate a high-speed train running between Windsor and Quebec City to the provinces of Ontario and Quebec. The federal government would delegate authority over the Ontario operations of the undertaking to an Ontario board, while the Quebec operations would be subject to the authority of a Quebec board. The high-speed rail operation would then be subject to the authority of two separate regulators, one established by Ontario, the other by Quebec.

There would seem to be no way to avoid such divided jurisdiction, at least under the terms of the existing Constitution. It does not appear to be possible for the federal government to delegate authority over the complete operations of an interprovincial undertaking to a single provincially appointed board since no single province could validly create a board with such a mandate. An individual province is permitted only to create a tribunal with authority to regulate transportation undertakings operating within that province. For the same reason, no combination of provinces, acting in concert, can create a single board with authority to regulate the complete operations of an interprovincial transportation undertaking. Thus any attempt by the federal government to delegate regulatory powers over a complete interprovincial undertaking to a single provincially appointed

board would be unconstitutional. This would amount to an attempt by the federal government to enlarge the legislative powers of the provinces, a form of delegation which is not permitted under the current constitution.¹¹²

If this analysis is correct, any delegation of regulatory authority over interprovincial undertakings such as ferries or railways would inevitably create a situation of divided regulatory authority. The major drawbacks to any delegation of this type could only be avoided through an amendment of the Constitution to provide for direct delegations of legislative power between the two levels of government. Significantly, the current federal proposals contemplate an amendment precisely along these lines. The federal government proposes to insert provisions in the Constitution "to enable the delegation of legislative authority between the two levels of government with the mutual consent of the legislative bodies involved."¹¹³ If such provisions were enacted, they would permit the federal government to delegate authority over the complete operations of an interprovincial undertaking to a single provincially appointed tribunal.

There is a long history of similar proposals to permit delegations of powers directly between governments. The Fulton-Favreau amendment proposal of 1964 would have inserted a power of interdelegation in the Constitution. More recently, the Macdonald Commission recommended a constitutional amendment to permit legislative as well as administrative delegation of powers.¹¹⁴ A similar proposal was endorsed by the Beaudoin-Edwards Committee examining the constitutional amending formula.¹¹⁵

Proposals along these lines have certainly attracted some measure of criticism in the past.¹¹⁶ On balance, however, a carefully framed interdelegation power would appear to represent a positive contribution to our existing constitutional framework.¹¹⁷ Such a power would permit greater flexibility in the way in which governments respond to social problems. In transportation, for example, it would open the door to a greater range of regulatory responses involving both levels of government. In particular, it would allow the federal Parliament to allocate responsibility over transportation matters on a regional as opposed to a purely provincial basis. There might well be circumstances, such as a high-speed rail service in the Windsor-Quebec City corridor, where such a regional form of regulation is appropriate. A constitutional amendment permitting legislative delegation of powers between the two orders of government would make this form of regional, interprovincial responsibility for transportation matters a possibility.

Such delegation of powers might also make it possible to fashion more creative responses to the difficulties associated with divided jurisdiction over transportation matters. As noted in the previous section, actions and policies of one level of government often significantly affect the activities of other levels of government. Furthermore, competition within the transportation field occurs across transportation modes as well as within a particular mode. If governments could delegate jurisdiction directly, it might be possible to create new regulatory structures to respond to these challenges more effectively. For example, short air trips compete directly with trips by cars and trains. It might be desirable to grant a single level of government the authority to regulate all three modes of transportation for trips involving limited distances.¹¹⁸ Permitting direct delegations of legislative powers would add greater flexibility in the responses that are available to government. It would ensure that government regulation can be more closely tailored to the realities of the transportation industry as it evolves.

IV. CONCLUSION

This paper has reviewed the current allocation of jurisdiction over transportation matters and assessed proposals for constitutional change. These arrangements have worked fairly well in the past because of the pragmatic interpretation of the division of powers by the courts. In contrast to their approach to many other areas of the Constitution, the courts have tended to employ a functional analysis in construing constitutional responsibilities over transportation matters. They have rejected the idea of dividing jurisdiction over a single undertaking between two levels of government. They have also held that where two transportation undertakings are functionally integrated or operated in common, they should be subject to a single regulatory authority. The practical effect of this approach has been to expand federal authority in this field. Since jurisdiction over particular undertakings is to be undivided, federal authority has been recognized over undertakings that are involved only minimally in interprovincial transportation. Similarly, federal authority has also been recognized in relation to local transportation undertakings which are functionally connected or integrated with interprovincial undertakings.

Because of this expansive reading of federal power, the Constitution has made it possible for a single level of government to play a leading role in transportation policy. The federal government has been able to facilitate

the creation of a truly national transportation network across the country. I conclude that there is no need for major constitutional changes in the existing jurisdiction over transportation matters in the Canadian Constitution. The only only significant change which I propose is an amendment to permit the direct delegation of legislative powers between governments. This proposal, advanced recently by the federal government, would provide for greater flexibility and more effective coordination of the respective roles of governments. It would permit governments to adjust their responsibilities on an incremental basis in response to changing trends and circumstances in the transportation industry. In this way, the Constitution would not block or impede the effort to adapt our transportation system to the needs of future generations of Canadians.

ENDNOTES

1. The focus of this study is on the provisions which grant legislative authority in the field of transportation. A companion study in this volume by Patrick J. Monahan, entitled "Transportation Obligations and the Canadian Constitution," examines a number of constitutional provisions which impose a particular obligation to provide transportation services. The constitutional obligations of the federal government will not be considered in any detail in the present study.
2. Government of Canada, *Shaping Canada's Future Together: Proposals* (Ottawa: Supply and Services Canada, 1991).
3. See, for example, *A Quebec Free to Choose, Report of the Constitutional Committee of the Quebec Liberal Party* (Montreal, January 1991); *Some Practical Suggestions for Canada*, Report of the Group of 22 (Montreal, 1991).
4. See, for example, C. McNairn, "Transportation, Communication and the Constitution," 47 *Canadian Bar Review* 355 (1969); C. McNairn, "Aeronautics and the Constitution," 49 *Canadian Bar Review* 411 (1971); C. Dalfen and L. Dunbar, "Transportation and Communications: The Constitution and the Canadian Economic Union," in *Case Studies in the Division of Powers*, ed. M. Krasnick, (Royal Commission on the Economic Union, Background Studies, Vol. 62, 1985), pp. 139–202.
5. For example, the extent to which the development of so-called "intermodal" forms of transportation might be constrained or limited by the Constitution is examined. Intermodal transportation involves the integration of different modes of transportation within a single network or system. The integrated system would connect modes either through a single reservation system or through common facilities, such as train stations, airports and bus terminals. See *Getting There: The Interim Report of the Royal Commission on National Passenger Transportation* (Ottawa: Supply and Services Canada, April 1991), p. 141. It is important to emphasize that this aspect of the study focusses on constraints which are implicit in the scheme of the Constitution itself, as opposed to government regulation or legislation. An examination of the efficacy or limitations of existing government policy in the transportation field is beyond the scope of the present study.
6. See *A Quebec Free to Choose*.

7. See *Some Practical Suggestions for Canada*.
8. See *Shaping Canada's Future Together: Proposals*.
9. The issue here is the extent to which regulatory authority can be transferred from one level of government to another without the need to resort to a formal constitutional amendment.
10. See Peter Hogg, "Transportation and Communication," in *Constitutional Law of Canada*, 2nd ed., (Toronto: Carswell, 1985), p. 484.
11. Section 92(10)(a), *Constitution Act, 1867*, U.K., 30, 31 Victoria, c. 3.
12. See the Third Schedule to the *Constitution Act, 1867*, which sets out the classes of provincial property which passed to the Dominion at the time of Confederation. The transfer of title to the Dominion was effective at the moment of each province's entry into Confederation; as transportation was in a stage of infancy at the relevant dates, Dominion proprietary rights acquired in this way have not proven to be significant factors in the regulation of transportation. See McNairn, "Transportation, Communication and the Constitution," p. 366.
13. It should be noted, however, that the trade and commerce power has been used to regulate the safety of new cars and components. The *Motor Vehicle Safety Act*, R.S.C. 1985, c. M-10 requires all motor vehicles imported into Canada to comply with federal safety and environmental regulations. In addition, vehicles manufactured in Canada must have a National Safety Mark indicating that they meet the relevant federal safety and environmental standards.
14. See the *Criminal Code* of Canada, R.S.C. 1985, c.C-46, sections 249–261.
15. *Re Regulation and Control of Radio Communication in Canada* [1932] A.C. 304, p. 315.
16. See *A.G. Ontario v. Winner* [1954] A.C. 541.
17. See *Toronto v. Bell Telephone* [1905] A.C. 52.
18. *A.G. B.C. v. A.G. Can. (Natural Products Marketing Reference)* [1937] A.C. 377.
19. It should be noted that in recent years, the Supreme Court of Canada has moved away from this bifurcated approach to the trade and commerce power, recognizing federal authority to regulate general trade and commerce. See *General Motors of Canada Ltd. v. City National Leasing* (1989), 58 D.L.R. (4th) 255 (S.C.C.).
20. Of course, while this constitutional capacity has been vested in the federal government, it has not necessarily chosen to exercise the jurisdiction to its limits. See, for example, the *Motor Vehicle Transport Act*, 1987, R.S.C. 1985, c. 29 (3rd Supp.) s.4 (delegating authority to regulate extra-provincial bus undertakings to the provinces).
21. *Alberta Government Telephones v. Canadian Radio-television and Telecommunications Commission and CNCP Telecommunications* [1989] 2 S.C.R. 225, p. 257 (per Dickson, C. J.).
22. *Re Tank Truck Transportation Ltd.* (1960), 25 D.L.R. (2d) 161, Ont. High Court.
23. This result was affirmed by the Court of Appeal without written reasons. See [1963] 1 O.R. 272 (C.A.).

24. *R. v. Cooksville Magistrate's Court, Ex parte Liquid Cargo Lines Ltd.* [1965] 1 O.R. 84.
25. *AGT v. CRTC* [1989] 2 S.C.R. 225.
26. All the members of the Court agreed on the issue of jurisdiction; Madame Justice Wilson dissented on the issue of whether AGT was entitled to assert a claim of Crown immunity.
27. *Ibid.*, p. 259.
28. *Ibid.*, p. 257.
29. *Ibid.*, p. 258.
30. *Ibid.*, p. 262.
31. See P. Hogg, "Comments," (1990) 35 *McGill L. J.*, p. 480.
32. *Re Cannet Freight Cartage* [1976] 1 F.C. 174 (C.A.) (approved by the Supreme Court of Canada in *UTU v. Central Western Railway* [1990] 3 S.C.R. 1112, 1145-47.)
33. See Hogg, "Comment," p. 487.
34. Federal authority over Bell Canada and B.C. Tel was based on provisions in their respective Special Acts declaring them to be works "for the general advantage of Canada"; telephone service in Yukon, NWT and Newfoundland is provided by subsidiaries of CN Railway, which is subject to federal authority by virtue of provisions in the *Canadian National Railways Act*.
35. For a summary and discussion of the regulatory situation prior to the AGT case, see Dalfen and Dunbar, pp. 156-160.
36. Because the Supreme Court also held that AGT was entitled to benefit from the doctrine of Crown immunity, the result of the case was that the telephone company was subject neither to federal nor to provincial authority. The other two prairie telephone companies in Saskatchewan and Manitoba were in a similar situation. However, the four Atlantic telephone companies are privately owned and can claim no Crown immunity; the effect of the Court's decision was to immediately bring the four Atlantic companies within federal authority. In October of 1989 the federal government introduced Bill C-41, an amendment to the *Railway Act* to make the Act binding on the three prairie telephone companies. However, this bill died on the order paper and has not been reintroduced (as of October 25, 1991). At the moment, therefore, there is no federal legislation which is applicable to the three prairie telephone companies.
37. See *Montreal v. Montreal Street Railway* [1912] A.C. 333.
38. *Luscar Collieries Limited v. McDonald* [1927] A.C. 925.
39. *Reference re Industrial Relations and Disputes Investigation Act* [1955] S.C.R. 529.
40. *Letter Carriers' Union of Canada v. Can. Union of Postal Workers* [1975] 1 S.C.R. 178.
41. *UTU v. Central Western Railway* [1990] 3 S.C.R. 1112; (1990) 76 D.L.R. (4th) 1.

42. Dickson C.J., spoke for eight members of the Court on the constitutional issue; Madame Justice Wilson was the sole dissenter.
43. *UTU v. CWR*, 76 D.L.R. (4th), pp. 15–16.
44. *Ibid.*, pp. 19–21.
45. It was argued that there was an integrated network for the transportation of grain within Western Canada and that Central Western formed part of this network. However, Dickson C. J. concluded that “I do not agree that a Western Grain Transportation Network exists for the purposes of the jurisdictional designation of the Central Western. . . . [T]he fact that several entities involved in the transport of grain fall under federal jurisdiction cannot on its own serve to bring everything connected with that industry under federal jurisdiction.” *Ibid.*, p. 22.
46. See Hogg, “Transportation and Communication,” p. 491.
47. See *Jorgensen v. A.G.Can.* [1971] S.C.R. 725.
48. *Ibid.*
49. See N. Finkelstein, *Laskin’s Canadian Constitutional Law*, 5th ed. (Toronto: Carswell, 1986), pp. 627–631.
50. See A. Lajoie, *Le pouvoir déclaratoire du Parlement* (Montreal: University of Montreal, 1969) pp. 70–72.
51. See *Cape Breton Development Corporation Act*, S.C. 1967, c.6, s.35(1); *Teleglobe Canada Reorganization and Divestiture Act*, S.C. 1987, c.12, s.9.
52. See *Shaping Canada’s Future Together: Proposals*, proposal no. 23.
53. See *Johannesson v. West St. Paul* [1952] 1 S.C.R. 292.
54. *Ibid.*, p. 308–09.
55. See *Jorgenson v. North Vancouver Magistrates* (1959) 28 W.W.R. 265 (B.C.C.A.).
56. See *Re Orangeville Airport Ltd.* (1976) 11 O.R. (2d) 546 (C.A.).
57. See *Re Walker and Minister of Housing for Ontario* (1983) 41 O.R. (2d) 9 (C.A.).
58. See *Field Aviation Company Limited v. Alberta Board of Industrial Relations* [1974] 6 W.W.R. 596 (Alta. A.D.).
59. See *Construction Montcalm Inc. v. The Minimum Wage Commission* [1979] 1 S.C.R. 754.
60. See *Re Colonial Coach Lines Ltd.* [1967] 2 O.R. 25 (Ontario H.C.); *Murray Hill Limousine Service v. Batson* [1965] B.R. 778 (Que. C.A.).
61. In addition to the three classes of subject noted in the text, the power over “trade and commerce” (section 91.2) and the power over “criminal law” (section 91.27) have been used to enact legislation that relates to transportation policy. See discussion *supra* at notes 13 and 14.

62. See *Attorney General of Canada v. Attorney General of Ontario, Quebec and Nova Scotia*, [1898] A.C. 700.
63. This, for example, was the view expressed by Dalfen and Dunbar in their study for the Macdonald Commission in 1985; see Dalfen and Dunbar, p. 150.
64. *Agence Maritime Inc. v. Canada Labour Relations Board*, [1969] S.C.R. 851.
65. *Singbeil v. Hansen* (1985) 19 D.L.R. (4th) 48 (B.C.C.A.).
66. *Whitbread v. Walley* (1990) 120 N.R. 109.
67. *Ibid.*, p. 120.
68. *Ibid.*, p. 123.
69. For example, labour relations matters in all aeronautics undertakings are subject to exclusive federal authority; in the field of navigation and shipping, federal authority has been limited to undertakings engaged in interprovincial and international activity.
70. See discussion *supra* at note 18 and accompanying text.
71. For example, the regulatory framework with respect to the road system or to private passenger automobiles is not included in this discussion. Neither of these areas has been the source of any significant litigation or court decisions.
72. For an extended discussion of the provincial role in this regard see IBI Group, "Intercity Passenger Transportation Policy Framework: Provincial Economic and Safety Legislation Review" a working paper prepared for the Royal Commission on National Passenger Transportation, October 1990, pp. 3–7.
73. *Ibid.*, p.4. Overall provincial involvement in operating and subsidizing airports has been increasing in the past decade: in 1979, the provinces operated or subsidized about 124 land airports, whereas in 1989 this number had grown to 378. However, Transport Canada continues to operate the largest and busiest airports in Canada and its annual airport related expenditures are about 11 times higher than those of the provinces combined.
74. The province of Ontario has established norOntair to provide subsidized air services to 21 communities in Northern Ontario. In 1988, the operating subsidy for norOntair was \$4.3 million.
75. Certain of these ferry services are a constitutional obligation of the federal government, including ferries in Prince Edward Island, Newfoundland and British Columbia. These obligations are contained in the respective Terms of Union admitting these provinces to Canada. Each of the obligations are thus part of the Constitution of Canada and take precedence over all federal and provincial legislation. Further, they may only be amended in accordance with the procedures set down in Part V of the *Constitution Act, 1982*. For a discussion of the precise nature and implications of these various obligations, see Monahan, "Transportation Obligations and the Canadian Constitution."
76. The federal government has, in recent years, sought to devolve responsibility for the provision of intra-provincial ferry services to the provinces. The usual arrangement has the province agreeing to assume responsibility for a service in exchange for payment of a fixed sum from the federal government. For a discussion, see IBI Group, "Intercity

Passenger Transportation Policy Framework: Federal Legislation Review," a working paper prepared for the Royal Commission on National Passenger Transportation, June 1990, pp. 14–16.

77. For example, British Columbia has established a Crown corporation (British Columbia Ferry Corporation) to provide ferry services in the province. Tolls and service schedules are established by the Corporation's board of directors and must be approved by the provincial Cabinet. The Corporation also receives a substantial subsidy from the provincial government.

In Newfoundland, the province directly operates seven intra-provincial coastal ferry services, while an eighth service is subsidized. These services were taken over from the federal government in 1979. Service schedules and fares are developed by the Newfoundland Department of Transportation.

The Quebec Ministry of Transportation operates or subsidizes about a dozen longer-distance ferry routes across the St. Lawrence River and along the North Shore of the Gulf of St. Lawrence.

For further descriptions of these various provincially operated or subsidized services, see IBI Group, "Intercity Passenger Transportation Policy Framework: Provincial Economic and Safety Legislation Review," pp. 14–19.

78. The British Columbia Railway provides passenger service between North Vancouver and Prince George, B.C.
79. The Ontario Northland Railway offers passenger rail service in Northern Ontario, while GO Transit provides mainly commuter rail services, some of them on its own trackage, in the Greater Toronto Area. Ontario has also established the Ontario Northland Transportation Commission and granted it the authority to operate and regulate rail services in Northern Ontario. Decisions of the Commission are subject to amendment by the provincial Cabinet.
80. The Montreal Urban Community Transport Commission offers commuter rail services in the Montreal area.
81. This principle was established by the Supreme Court of Canada in *A.G. Nova Scotia v. A.G. Canada* [1951] S.C.R. 31.
82. *Coughlin v. Ontario Highway Transport Board* (1968), 68 D.L.R. (2d) 384.
83. For a discussion of these provincial variations, see IBI Group, "Intercity Passenger Transportation Policy Framework: Provincial Economic and Safety Legislation Review," pp. 20–33.
84. See section 16 of the MVTA, 1987 which provides as follows:

The Governor in Council may, by regulation, on the recommendation of the Minister made after consultation by the Minister with the government of each province affected thereby, exempt from the application of this Act or of any provision of this Act, either generally or for a limited period or in respect of a limited area, any person, the whole or any part of any extra-provincial bus undertaking or extra-provincial truck undertaking, every extra-provincial

bus undertaking or extra-provincial truck undertaking, any group or class of such undertakings or any extra-provincial bus transport or extra-provincial truck transport.

85. The Roadcruiser bus service, operated by CN in Newfoundland, was initiated in 1968 when CN shut down the passenger rail service in Newfoundland and substituted a bus service in its place. It was exempted from the MVTA in the mid-1970s following a dispute between CN and the Board of Commissioners of Public Utilities of Newfoundland.
86. This is the view expressed by Peter Hogg in his discussion of the *Coughlin* case. See Hogg, *Constitutional Law of Canada*, 2nd ed., p. 303.
87. This limitation will prove to be significant in our discussion later in this paper.
88. See *Getting There*, pp. 141–42.
89. *Ibid.*, p. 40.
90. See, for example, the *Canadian Transportation Accident Investigation and Safety Board Act*, S.C. 1989, c.3.
91. See *Getting There*, p. 44.
92. See the provisions of the *Motor Vehicle Transport Act, 1987* noted in the previous section, delegating authority over the bus industry to the provinces.
93. *Canadian Transportation Accident Investigation and Safety Board Act*, S.C. 1989, c.3. See section 3 for a definition of the application of the Act.
94. See the *Motor Vehicle Safety Act*, R.S.C. 1985, c.M-10.
95. See discussion in *Getting There*, pp. 103–05.
96. The possibility of establishing some sort of Canadian Terminals Agency is discussed in *Getting There*, p. 149.
97. For a discussion of these various factors, see *Getting There*, pp. 104–05.
98. *Ibid.*, p. 226.
99. See *Some Practical Suggestions for Canada*.
100. *A Quebec Free To Choose*.
101. *Shaping Canada's Future Together*, p. 58.
102. *Ibid.*, see pp. 58–59.
103. See Hogg, *Constitutional Law of Canada*, p. 493.
104. See *Atomic Energy Control Act*, S.C. 1946, c.37.

105. This was the reasoning of the Ontario High Court in *Pronto Uranium Mines Limited v. The Ontario Labour Relations Board*, [1956] O.R. 862.
106. Hogg (*Constitutional Law of Canada*, p. 493) suggests that the declaratory power has not been used since 1961.
107. See the *Telegraph Canada Reorganization and Divestiture Act*, S.C. 1987, c.12, s.9.
108. See, for example, the *Cape Breton Development Corporation Act*, S.C. 1967, c.6, which recites that the provincial government had consented to the legislation containing a federal declaration under section 92(10)(c) of the *Constitution Act, 1867*.
109. These suggestions have been made before and are reviewed and discussed in Hogg, *Constitutional Law of Canada*, p. 493.
110. See *Shaping Canada's Future Together*, Proposal 22.
111. This was the basis of the reasoning by the Supreme Court of Canada in the *Johannesson* case, *supra* note 53.
112. This is the basic proposition established by the *Nova Scotia Interdelegation Case*. For a discussion of the principles established by this case, and an analysis which leads to similar conclusions proposed in the text, see Hogg, *Constitutional Law of Canada*, pp. 307–08.
113. See *Shaping Canada's Future Together*, Proposal 25.
114. See The Macdonald Commission, *Report of the Royal Commission on the Economic Union and Development Prospects for Canada*, Vol. III, p. 257.
115. See Special Joint Committee of the Senate and the House of Commons, *The Process for Amending the Constitution of Canada*, June 20, 1991, p. 29.
116. Critics of legislative interdelegation argue that it would promote uncertainty and would "confuse the basic political responsibility and accountability of members of the federal Parliament and the federal Cabinet, and too much of this could destroy these federal institutions." See W. R. Lederman, "Some Forms and Limitations of Co-operative Federalism," 45 *Canadian Bar Review* 409 (1967), p. 426.
117. The interdelegation power would have to be framed so as to ensure that a delegation might be revoked upon the giving of proper notice. The delegation would also have to be framed in relatively precise terms and in accordance with certain principles or criteria agreed upon in advance by both levels of government.
118. This is suggested in *Getting There*, p. 227.

TRANSPORTATION OBLIGATIONS AND THE CANADIAN CONSTITUTION

Patrick J. Monahan*

October 1991

I. INTRODUCTION: TRANSPORTATION AND THE CANADIAN CONSTITUTION

The problems and methods of transportation constitute an essential thread in the development of the country.¹

Transportation policy has always occupied a critical and central place in Canadian nation-building. The close connection between transportation policy and national policy is nowhere more evident than in the terms of the Canadian Constitution itself. Canada is unique among the developed nations of the world for the number and detail of transportation obligations which have been entrenched in its formal constitution.

The *British North America Act, 1867* (now called the *Constitution Act, 1867*) recites the undertaking of the Canadian government to secure the construction of a railway linking the former colonies with each other; British Columbia entered the Union in 1871 in return for a constitutionally entrenched guarantee of a transcontinental railway; the *Prince Edward Island Terms of Union* require the Canadian government to maintain a ferry service linking the Island with the mainland; the *Newfoundland Terms of Union* (*Newfoundland Act*) provide a guarantee of ferry service between the new province and the

* Director, York University Centre for Public Law and Public Policy.

Canadian mainland. These are merely illustrations of a considerable list of similar constitutional provisions guaranteeing particular transportation services or infrastructure.

Students of Canadian history have long remarked on the number and specificity of these constitutional obligations.² The most significant of these relate to undertakings to construct or to take over railways in various parts of the country.³ It seems that the Canadian government fulfilled these obligations to the ultimate satisfaction of all concerned in the latter half of the 19th century; thus, while of considerable historical interest, such obligations have not been seen as playing a significant role in shaping modern transportation policy in Canada.⁴

More recently, however, interest in the status and meaning of these constitutional obligations has been revived because of the termination or reduction of passenger rail service in many parts of the country. Two provinces have argued that certain proposed reductions are unconstitutional on the basis that they violate constitutional guarantees given to these provinces when they joined Confederation. While Prince Edward Island's legal challenge failed,⁵ the challenge in British Columbia succeeded in the trial division of the British Columbia Supreme Court, with the Court ordering the reinstatement of passenger service on a rail line on Vancouver Island.⁶ This ruling has recently been upheld by the British Columbia Court of Appeal.⁷ The success of this challenge has raised the issue of whether this and other transportation obligations in the Canadian Constitution might indeed have a role to play in the future evolution of Canadian transportation policy.

This paper presents a comprehensive analysis of the legal status, meaning and significance of transportation obligations in the Canadian Constitution.⁸ It describes and analyzes the current legal status of the obligations to determine the extent to which they will influence, constrain or shape future transportation policy. While the focus of the paper is on the contemporary legal significance of the obligations, much of the analysis is historical because the current significance of the obligations can only be understood through a review of the purpose and meaning of the obligations when they were first enacted. The report details the circumstances surrounding the enactment of these various obligations as well as the then-prevailing understanding of their meaning and purpose. It also considers the manner in which the obligations have been carried out over the years and the extent

to which the original understanding of their meaning has been reflected in subsequent government policy or judicial decision. Finally, it assesses the current significance of the obligations and whether they should be considered by transportation policy makers concerned with meeting Canada's transportation needs into the next century.

The Canadian practice of specifying certain transportation obligations in its fundamental constitutional law has often been regarded as somewhat out of the ordinary. The larger question, however, is whether there is any reason in principle to object to the practice. The concluding section of this paper reflects on the wisdom of the Canadian approach to constitution making. Alternative methods are suggested which would stop short of formal constitutional entrenchment of an obligation to provide named transportation services.

II. THE CONFEDERATION ERA 1867-1873: TRANSPORTATION UNDERTAKINGS AS THE INSTRUMENT OF POLITICAL UNION

A. *BRITISH NORTH AMERICA ACT, 1867*

A variety of factors led to the union of the British North American provinces in the mid-1860s, including the fear of annexation by the United States as well as the termination of the reciprocity treaty by the Americans in 1865.⁹ But the key factor in securing support for the scheme, at least among the Maritime colonial leadership, was the commitment to construct an intercolonial railway. As one commentator has put it: "[c]ertainly there could have been in 1867 no confederation without the Intercolonial: there might have been an Intercolonial without confederation."¹⁰

By the early 1860s, the idea of constructing an intercolonial railway linking the Maritime provinces with Canada had been under discussion for at least two decades. Negotiations to build a rail link had been pursued actively but the discussions had ultimately foundered when the Canadians were unwilling to accept certain conditions demanded by Great Britain. By 1862, negotiations reached a stalemate.

When delegates from the various colonies of British North America gathered at Charlottetown in 1864 to discuss political union, the representatives from

New Brunswick and Nova Scotia saw the meeting as an opportunity to pursue their goal of securing a rail link with the markets of Canada. The Maritime delegates insisted that the construction of an intercolonial railway was a non-negotiable condition of their support for political union with Canada.¹¹ Nor were the Maritimers willing to accept a mere political commitment from the Canadians that such a railway would be constructed following Confederation. Instead, they insisted that the guarantee should be written into the terms of the imperial statute creating the new federation. Invoking the authority of Westminster would provide an ironclad guarantee that the Canadians would keep their promise to build the railway.

At the time there did not appear to have been any great controversy or objection to this way of explicitly guaranteeing the construction of the Intercolonial. Subsequent commentators have suggested that it is somewhat out of the ordinary to make explicit reference to the construction of a railway in a country's constitution. It should be remembered, however, that what was being contemplated was the enactment of an ordinary British statute. It is perfectly commonplace for ordinary legislation to make reference to quite specific matters, including the carrying out of contractual obligations. Clearly, the colonial leaders of British North America in the 1860s approached the matter on this footing. They were seeking to ensure that the commitment to build the Intercolonial could not be reversed by a subsequent Canadian government. The easiest and most straightforward way to secure this commitment was to set it out explicitly in an imperial statute. The British authorities would act as the "guardians" of the commitment, since any change in the terms of the undertaking would require the consent of Westminster.

The draft resolutions agreed to at the Quebec Conference in the fall of 1864 committed the new Canadian government to building the Intercolonial. Resolution 68 from the Quebec Conference reads as follows:

68. The General Government shall secure without delay the completion of the Intercolonial Railway from Rivière-du-Loup through New Brunswick to Truro in Nova Scotia.¹²

The importance attached to the construction of the Intercolonial is evident in a comparison of resolution 68 with the terms of resolution 69 from the Quebec Conference. Resolution 69, in its reference to the construction of

a railway linking the western territory and the Pacific Ocean with the proposed federation, offered the following commitment:

69. The communications with the Northwestern Territory and the improvements required for the development of the trade of the Great West with the Seaboard, are regarded by this Conference as subjects of the highest importance to the Federated Provinces, and shall be prosecuted at the earliest possible period that the state of the finances will permit.¹³

The delegates clearly thought that cost considerations would govern the timing of the construction of the rail link with the west. The commitment to construct the western railway, while regarded as a subject of the “highest importance,” was left to the discretion of the new government of the Dominion “at the earliest possible period that the state of Finances will permit.” There was no such concern for cost with respect to the Intercolonial. The construction of that railway was expressed in mandatory terms: the general government “shall” secure the completion of the Intercolonial “without delay.” The commitment was unqualified and unavoidable. The Canadian government was to secure construction of the railway without regard to considerations of cost or feasibility. The resolutions left no doubt that the construction of the Intercolonial was, from the point of view of the Maritimes, a “condition precedent” for political union.

The London resolutions in 1866 carried forward the explicit and unqualified commitment to construct the Intercolonial. Resolution 65 provided as follows:

65. The construction of the Intercolonial Railway being essential to the consolidation of the Union of British North America, and to the assent of the Maritime Provinces thereto, it is agreed that provision be made for its immediate construction by the General Government, and that the Imperial guarantee for three millions of pounds sterling pledged for this work be applied thereto, so soon as the necessary authority has been obtained from the Imperial Parliament.

The London version differed in certain important respects from the terms set out in the Quebec resolutions of 1864. First, the proposed railway was simply described as the “Intercolonial,” without any reference to the starting or endpoint of the line; the reference to Rivière-du-Loup and to

Truro had been dropped. Second, while the London version referred to the fact that the construction of the Intercolonial was "essential to the consolidation" of the British Colonies, the language expressing the commitment was surprisingly vague. Whereas the Quebec resolution had indicated in mandatory terms that the General Government "shall secure, without delay, the completion of the Intercolonial Railway," the London version simply referred to the fact that "provision be made for its immediate construction by the General Government."

Finally, the London version stated that the Imperial Government would guarantee a loan of three million pounds sterling to enable the construction to proceed. The addition of this guarantee was of great practical value and represented the most significant modification from the earlier Quebec resolutions. Constitutional commitments are hardly worth the paper they are written on if they are not backed by the necessary funds. The financial guarantee from Great Britain would ensure that the constitutional commitment to build the Intercolonial would become a practical reality.

The wording of the London resolutions served as the basis for the drafting of the *British North America Act* (BNA Act) in January and February of 1867.¹⁴ Section 145 of the *British North America Act, 1867*, as enacted by the British Parliament, set out the formal commitment to build the Intercolonial in the following terms:

145. Inasmuch as the Provinces of Canada, Nova Scotia, and New Brunswick have joined in a Declaration that the Construction of the Intercolonial Railway is essential to the Consolidation of the Union of British North America, and to the Assent thereto of Nova Scotia and New Brunswick, and have consequently agreed that Provision should be made for its immediate Construction by the Government of Canada: Therefore, in order to give effect to that Agreement, it shall be the Duty of the Government and Parliament of Canada to provide for the Commencement, within Six Months after the Union, of a Railway connecting the River St. Lawrence with the City of Halifax in Nova Scotia, and for the Construction thereof without Intermission, and the Completion thereof with all practicable Speed.

It can be seen that the first half of section 145 tracks precisely the terms of the comparable London resolution. It recites the agreement that the construction of the Intercolonial was essential to the union and that "provision

should be made for its immediate Construction.” The second half of the section, however, is new, setting out a “Duty” of the Government and Parliament of Canada. Certain features of the duty should be emphasized:

1. Whereas previous versions had referred only to the obligation of government alone, section 145 established a duty of the Government and *Parliament* of Canada. This is significant, since the duty to build the railway was thereby entrenched as a limitation on the legislative authority of the Canadian Parliament itself.
2. The obligation had a specific time frame attached; the construction was to commence within six months and was to proceed “without Intermission and . . . With all practicable Speed.” Significantly, however, there was no specific date set for completion of the railway.¹⁵
3. The Intercolonial is specifically identified as linking the “River St. Lawrence with the City of Halifax,” similar to the wording of the original Quebec resolution in 1864.
4. Section 145 makes no reference to the financial guarantee of the imperial government. This was because the financial guarantee was secured by a second imperial statute passed contemporaneously, entitled *An Act for authorizing a Guarantee of Interest on a Loan to be raised by Canada towards the Construction of a Railway connecting Quebec and Halifax*. Under this bill, three million pounds were provided for the construction of the railway.¹⁶

While the Parliament and Government of Canada undertook to secure the construction of the railway, section 145 is silent as to:

- the route to be selected for the railway;
- the amount to be spent on its construction; and
- the manner in which the railway would be operated once it has been completed, including such matters as fares and the general level and quality of service on the road.

However, these matters were dealt with elsewhere in the Act. Sections 92(10)(a) and 91(29) of the BNA Act granted the federal Parliament exclusive legislative authority over interprovincial railways. Thus the Parliament of Canada,

as opposed to the legislatures of the provinces, was granted exclusive legislative authority over the Intercolonial Railway. While Parliament's legislative authority was fettered to the extent that it was obliged to construct the railway, upon completion Parliament would be free to regulate the actual operation of the railway in accordance with the policy of the government of the day.

The larger significance of section 145 was in the precedent it established. Having secured the consent of New Brunswick and Nova Scotia to Confederation with the promise of a railway, the Canadian authorities would soon find that similar promises would be demanded from other prospective provinces. The Canadians would also find that the price would rise with the passage of time; the undertakings which Canada was asked to take on would become more costly and the terms and conditions more difficult to fulfil. Still, having once agreed to such a request, the Canadians could find little reason to reject subsequent proposals. The only issue to be negotiated was the extent and nature of the transportation obligations to be shouldered by the new federal government.

B. WESTERN EXPANSION: THE *MANITOBA ACT, 1870* AND THE *BRITISH COLUMBIA TERMS OF UNION (1871)*

The *British North America Act, 1867* made specific provision for the eventual expansion of the country westward to the Pacific Ocean.¹⁷ In 1870 and 1871, the new nation began to make good on this promise of an expanded union as the provinces of Manitoba and British Columbia entered Confederation. But the terms on which the two provinces joined Canada were quite different, at least in the transportation obligations which were constitutionally assumed by the Canadian government.

For Manitoba, there were no specific constitutional entitlements to transportation services or infrastructure included in the *Manitoba Act, 1870*. In fact, the Act stated that any provision setting out a particular obligation regarding another province or group of provinces had no application with respect to Manitoba.¹⁸ This meant, for example, that the guarantee of a transcontinental railway made to British Columbia was made to that province alone and had no application to Manitoba. The explanation for this variation in approaches is straightforward. Manitoba, unlike British Columbia, did not have to be persuaded to join Confederation. The huge land mass to

the west of Ontario known as Rupert's Land and the North-Western Territory was annexed to Canada by imperial Order in Council in 1870.¹⁹ Immediately following the admission of the territories, the federal Parliament by statute created the province of Manitoba out of part of Rupert's Land. Since the land in question was already included within the Dominion, there was no need to use transportation promises to induce the new province to join Confederation.

Such was not the case with British Columbia. In 1867, the colony of British Columbia had watched the creation of the Canadian federation with great interest, even to the extent of passing a resolution in favour of admission to the Dominion. Following the annexation of Rupert's Land in 1870, delegates were sent from British Columbia to Ottawa to discuss the terms of B.C.'s entry to the federation.

British Columbia's interest in union with Canada did not prevent the western delegates from seeking advantageous terms for joining. Before leaving Victoria, the Legislative Council had debated and approved a series of specific proposed terms for admission to Canada, including specific undertakings for transportation services or infrastructure which were to be guaranteed by the Dominion. As in the case of the Intercolonial Railway, these were to be explicitly set out in the Terms of Union and thus permanently bind the Government of Canada. The most significant of these proposed guarantees was for a transportation link between British Columbia and the rest of the Dominion. British Columbia sought two commitments from Canada in this regard:

1. Construction of a coach road between B.C. and Fort Garry, to be completed within three years;
2. The immediate commencement of surveys for a transcontinental railway, with Canada committed to completion of the railway "at the earliest practicable date"; construction of the B.C. portion of the railway was to commence within three years.²⁰

Note that, while the Dominion was asked to actually construct the coach road and to open it for traffic within three years, no firm time frame was even requested with respect to the construction of a railway. The B.C. proposal simply asked that the railway be constructed "at the earliest practicable date"; what was "practicable" would no doubt be influenced

by considerations of cost and feasibility. At the time, no surveys had been made of the route through the mountains of British Columbia. Thus the overall cost of the undertaking was unknown and probably unknowable in advance.²¹ In short, the opening proposal from the British Columbians clearly contemplated that it might be quite some time before the trans-continental railway was actually completed.

What is somewhat surprising is that the final terms agreed upon went beyond the colony's request. Term 11 of the B.C. Terms of Union provided as follows:

11. The Government of the Dominion undertake to secure the commencement simultaneously, within two years from the date of the Union, of the construction of a railway from the Pacific towards the Rocky Mountains, and from such point as may be selected, east of the Rocky Mountains, towards the Pacific to connect the seaboard of British Columbia with the railway system of Canada; and further, to secure the completion of such railway within ten years from the date of the Union.

And the Government of British Columbia agrees to convey to the Dominion Government, in trust, to be appropriated in such manner as the Dominion Government may deem advisable in furtherance of the construction of the said railway, a similar extent of public lands along the line of railway throughout its entire length in British Columbia, not to exceed, however, twenty (20) miles on each side of said line, as may be appropriated for the same purpose by the Dominion Government from the public lands in the north-west territories and the Province of Manitoba. Provided that the quantity of land which may be held under pre-emption right or by Crown grant within the limits of the tract of land in British Columbia to be so conveyed to the Dominion Government shall be made good to the Dominion from contiguous public lands; and provided further, that until commencement, within two years, as aforesaid, from the date of union, of the construction of the said railway, the Government of British Columbia shall not sell or alienate any further portions of the public lands of British Columbia in any other way than under right of pre-emption, requiring actual residence of the pre-emptor on the land claimed by him. In consideration of the land to be so conveyed in aid of the construction of the said railway,

the Dominion Government agree to pay to British Columbia from the date of the Union, the sum of 100,000 dollars per annum, in half-yearly payments in advance.

The final version of the Terms of Union made no reference to British Columbia's original request for construction of a coach road linking the colony with Canada. However, the Government of the Dominion had assumed a much more onerous obligation regarding the construction of a transcontinental railway. The nature of the obligation can be discerned by comparing the main features of Term 11 with the original British Columbian request:

1. Term 11 stated that construction on the railway linking B.C. with the rail system of Canada had to begin *within two years* of the date of union and that construction had to start simultaneously in British Columbia and in Canada; under the original B.C. proposal, the only requirement was that construction of the initial sections of the railway *in B.C.* was to begin *within three years*;
2. Term 11 stated that the railway had to be completed *within 10 years* of the union; the original B.C. proposal merely required the construction to be completed at "the earliest practicable date."

It is unclear why the Canadians were prepared to accept such an onerous obligation, particularly the requirement that the railway be completed within 10 years. Early drafts of section 145 of the BNA Act had contemplated the establishment of a fixed completion date for the Intercolonial, but the Canadians resisted the idea that the BNA Act should specify a date for completion. In the B.C. case, the Canadians were prepared not only to fix a completion date for the railway but were prepared to do so for a project immeasurably more difficult than the Intercolonial.²²

The main explanation for this new approach appears to relate to the Canadian government's desire to bind its successors to its own railway policy.²³ Certainly the inclusion of the fixed completion date for the project aroused considerable opposition when it was debated in the Canadian Parliament. The leader of the Liberal opposition, Alexander Mackenzie, moved an amendment to the Terms of Union to the effect that Canada should be

pledged only to make surveys and to build the railway as finances might allow. But Mackenzie's amendment was defeated, and the unqualified obligation to complete the railway within 10 years was approved.

It should be noted that the formal constitutional obligation of Canada was limited to the actual construction of the railway; there is no mention of operation of the railway. Once the construction was completed, Canada would have fulfilled its obligation under Term 11. The absence of any requirement to "maintain" the railway after its completion was not accidental. In formulating its original request, the B.C. Legislative Council had instructed its negotiators to seek a commitment that Canada would construct "and maintain" a coach road linking the colony with Canada.²⁴ Similarly, B.C. had requested that an "efficient Coast Mail Service . . . be established and maintained. . . ." ²⁵ Thus the omission of reference to any requirement to "maintain" the railway suggests that the obligation was simply to construct the railway and not to operate or maintain it afterward.

There were other transportation obligations assumed by the federal government under the B.C. Terms of Union. Under Term 4, the Dominion undertook to "provide an efficient mail service, fortnightly, by steam communication between Victoria and San Francisco, and twice a week between Victoria and Olympia; the vessels to be adapted for the conveyance of freight and passengers." This obligation was in response to what had originally been two separate requests from the B.C. delegation: (a) that the Dominion government "supply an efficient and regular fortnightly steam communication between Victoria and San Francisco" and (b) that the Dominion government "establish and maintain" efficient coast mail service between Victoria, New Westminster, Nanaimo "and such other places as may require such services."²⁶ Steam communication between Victoria and San Francisco was requested because San Francisco was the western terminus for the American transcontinental railway, at the time, and the main means of communication between British Columbia and Canada. Ferry service to San Francisco then, was an important transportation link with the other Canadian provinces. At the same time, it was recognized that the San Francisco ferry service would become unnecessary once the Canadian transcontinental railway was completed.

During the debate over the Terms of Union in the B.C. Legislative Council, it was argued that the reference to a ferry service to San Francisco was a

mere “makeweight,” and not an essential condition of the colony’s entry into Confederation.²⁷ In the end, the British Columbian proposals requested that a ferry service to San Francisco be “supplied”; the comparable B.C. proposal dealing with the coastal mail service stated that such service was to be “established and maintained.” The absence of any requirement to “maintain” the San Francisco service might be taken to suggest that the obligation would not necessarily continue indefinitely.²⁸

Term 4 of the B.C. Terms of Union combines these two requests. Under Term 4, the Dominion is required to “provide” an efficient mail service linking Victoria with both San Francisco and Olympia on the mainland. The absence of the explicit requirement to “maintain” the service may indicate some ambiguity as to the duration of the commitment. There is no indication as to the precise meaning which the drafters associated with the requirement to “supply” mail service. Unlike Term 11 and the transcontinental railway however, Term 4 of the B.C. Terms of Union suggests that the obligation to supply ferry services in British Columbia was an obligation that would continue over time. Nor is there any indication that the obligation to “supply” the service would terminate at some fixed point in time. It seems that the Dominion requirement to “supply” the named ferry service was expected to continue indefinitely, until such time as the Terms of Union themselves were amended.²⁹

C. THE *PRINCE EDWARD ISLAND TERMS OF UNION* (1873)

While Prince Edward Island delegates had participated in both the Charlottetown and Quebec conferences in 1864, the Islanders were opposed to the proposed terms of Confederation as reflected in the Quebec resolutions. One of the main complaints was that the Quebec resolutions required construction of the Intercolonial; the P.E.I. delegates regarded the Intercolonial as imposing a heavy tax burden on Island residents without any corresponding benefit.³⁰ Following the Quebec Conference, the Prince Edward Island colonial Legislature passed resolutions rejecting the proposed terms of Confederation.

Following Confederation, negotiations to secure the Island’s admission to Canada continued. In December of 1869, the Canadian government made a formal offer of new terms that were a significant improvement over what had been offered to P.E.I. under the Quebec resolutions. One of the terms in

the Canadian offer was for provision of a ferry service linking the Island with the mainland. The Canadian offer stated that "Efficient Steam Service for the conveyance of mails and passengers was to be established and maintained between the Island and the Mainland of the Dominion, Winter and Summer, thus placing the Island in continuous communication with the International Railway and the Railway system of the Dominion."³¹ Although the 1869 offer had been framed in accordance with the advice of R.P. Haythorne, the Premier of the Island, it was still regarded as inadequate by the Island government.

One of the main objections was that the new terms made no provision for construction of a railway on the Island. At a series of public meetings held in Prince Edward Island in early 1870, it was argued that the terms would be acceptable only if they were supplemented by a sum sufficient for the construction of a railway. It was suggested that since P.E.I. would have to bear a portion of the expense of building Canadian railways, it was appropriate for the other provinces to contribute towards the construction of a railway on the Island.³² In April of 1870, the P.E.I. Legislative Council unanimously rejected the proposed Canadian terms.

In 1871, the Island decided to undertake construction of a railway on its own, without assistance from Canada. Legislation was passed providing for the construction of a railway the length of the Island, with the contractors to be paid through the issuance of debentures. The P.E.I. government estimated that the annual interest on the required capital would amount to 30,000 pounds, money that could be raised through a combination of modest tax increases and revenues raised from the operation of the road.³³ The actual costs, however, wildly exceeded these estimates. By 1873, total expenditures on the Island railway had exceeded \$3.2 million and the annual interest on the debt was nearly \$200,000. To place this debt load in perspective, the entire Island revenue from all sources for 1873 was a mere \$395,000.³⁴ The Island was simply unable to meet the financial obligations associated with the railway and looked to Confederation with Canada as a means of escaping bankruptcy.

Negotiations with the Canadian and imperial authorities began in early 1873. While the Canadians realized that the Island's financial crisis had prompted the negotiations, they agreed to relatively generous terms to secure the admission of P.E.I. to Confederation. The final Terms of Union were drafted

and mutually signed on May 15, 1873. The Dominion was to take over the railway that had threatened to cause the financial collapse of the Island government.³⁵ While the cost of the railway was to be charged against the Island as a local debt, the Dominion agreed to a “debt allowance” of \$50 per capita, twice the amount which had been agreed to in the case of the other provinces. P.E.I. was to receive an interest payment from Canada on the difference between its “debt allowance” and its actual debts upon admission.³⁶ Furthermore, since Canada had taken over the Island railway, all further costs associated with the railway would be the responsibility of the Dominion government rather than the Island.

It was further agreed that the Dominion government would “establish and maintain” a ferry service linking the Island with the mainland.³⁷ The use of the word “maintain” is significant; this was the first occasion that the Canadian government accepted the responsibility to ensure the continued operation of a transportation service. The undertaking was apparently open-ended, and there is specific reference to the fact that the ferry service must place the Island in “continuous communication” with the mainland. However, the use of the word “maintain” in connection with the ferry service coupled with its absence in the case of the Island railway is significant. It indicates that the obligation to “take over” the railway does not entail any obligation to “maintain” or actually operate the system.³⁸

One final transportation obligation of a relatively minor nature was included in the P.E.I. Terms of Union. The Dominion government assumed the cost, approximately \$2,500 per year,³⁹ of maintaining telegraph communication between the Island and the mainland.⁴⁰ As with the ferry service, the obligation to “maintain” the telegraph entailed an ongoing obligation until such time as the Terms of Union themselves were modified.⁴¹

III. MAKING GOOD ON THE OBLIGATIONS, 1873–1945:

TRANSPORTATION UNDERTAKINGS AS AN INSTRUMENT OF NATIONAL POLICY

Transportation undertakings played a key role in securing political support for Confederation among the colonial leadership in British North America. The distinctive feature of these undertakings was their explicit guarantee by imperial statute. Successive Canadian governments would be bound, as a

matter of law, to carry out the undertakings agreed to at the time of Confederation. The “guarantee” by Great Britain insulated them from the vagaries of Canadian party politics and from the unpredictable hazards of electoral contests. This was their overwhelming attraction to reluctant or doubtful political leaders in the various colonies.

It was one thing to assume certain obligations and have these obligations set out in imperial legislation. It was quite another matter actually to carry them out. The Intercolonial or the transcontinental railway may have existed on paper but passengers or freight could not be moved by means of constitutional language alone, no matter how majestic or unambiguous the wording. Moreover, the transportation undertakings which the Canadian government had assumed in the 1867–1873 period were extremely onerous, given the significant debt load which the new federation had inherited from the former colonies. There was no reason to assume that the government of the new Dominion would be capable of meeting all such obligations.

A. THE INTERCOLONIAL RAILWAY

In late 1867, the Parliament of Canada passed *An Act respecting the construction of the Intercolonial Railway*.⁴² This Act authorized the construction of the Intercolonial from Rivière-du-Loup to Halifax and placed the project under the authority of a board of four commissioners appointed by the Government (section 3). The Act also provided for a loan of four million pounds to finance construction, of which three million was guaranteed by the imperial government (sections 27, 32). The Chief Engineer, Sir Sandford Fleming, estimated the total cost of the project to be approximately \$20 million.

The legislation did not specify the route which the Intercolonial should follow. This was a matter of considerable controversy, particularly with respect to the location of the line through New Brunswick. Some argued that the line should run close to the American border in order to take advantage of railways which had already been constructed; others argued for a northern route which would be safer from a military point of view; still others suggested a central line through the province.

In the end, political and military considerations won out and the northern route was followed, a decision that proved to be a source of resentment and grievance for Maritimers for many years. Maritime political leaders charged

that the northern route added to the cost of construction while increasing the distance and the costs of transportation between Halifax and the markets of Central Canada. The decision in favour of the northern route, as well as the subsequent government management of the railway, contributed significantly to the sense of Maritimers' regional grievances over their status within the federation.⁴³

The railway was completed in 1876. The final cost of construction came to \$34 million, some \$14 million more than estimated, with the difference being financed entirely by the Government of Canada.⁴⁴ Control of the railway had been removed from the four commissioners and placed directly under the authority of the federal Minister of Public Works in 1874.⁴⁵ The Government of Canada found itself in the business of running a railroad.

From the day it opened, the Intercolonial was a losing proposition. In part, this was a product of the artificially low rates which were charged on the line. On average, freight rates on the Intercolonial were discounted approximately 20 percent from comparable rates in Central Canada. The Maritimes insisted that these discounted freight rates were appropriate, because of the selection of the longer northern route through New Brunswick. In 1927, this preferential rate structure for the Maritimes was given a statutory basis in the *Maritime Freight Rates Act, 1927* (S.C. 1927, c.44) whose purpose was described as to give "certain statutory advantages" in rates to persons and industries in "select territory" (section 7).

It must be emphasized that the decision to grant preferential rates was a matter of government policy rather than constitutional entitlement; nothing in section 145 of the BNA Act constrained the Government of Canada in its determination of rates on the Intercolonial. The lack of connection between section 145 and the rate structure is illustrated by the repeal of section 145 by the United Kingdom Parliament in 1893. Despite this repeal, the Maritimes insisted on continuation of the policy of discounted rates and the successive Canadian governments carried on with this preferential policy.

The Intercolonial never became profitable. In 1923, along with other government-owned railways (the P.E.I. Railway, the Grand Trunk and Grand Trunk Pacific and Canadian Northern), it was amalgamated into the Canadian National railway system.⁴⁶ Since then, what had been the Intercolonial has continued to be operated as a public enterprise by either CN or VIA Rail.

B. BUILDING THE CANADIAN PACIFIC RAILWAY

In 1871, the Canadian government decided that the railway to the Pacific would be built by a private corporation rather than by the government itself (see S.C. 1871, c.71). The government would provide assistance to the contractors through grants of land and subsidies, up to a limit of 50 million acres and \$30 million. But the negotiations between the government and potential contractors soon became mired in scandals and charges of corruption, eventually bringing down the Macdonald Government in 1873.⁴⁷

The subsequent Mackenzie administration had an entirely different attitude towards the construction of the Pacific railway, believing that the line could be built only as financial resources permitted. Prime Minister Mackenzie took the view that the promise of completion within 10 years was physically impossible to fulfil and should never have been given in the first place. Negotiations began between British Columbia and Ottawa in an effort to find some compromise acceptable to all parties.

The federal government sought an extension of the time for completion, and in return offered to construct a railway on Vancouver Island between Esquimalt and Nanaimo. The British Columbia government was unhappy with this proposal, maintaining that the construction of the Esquimalt-Nanaimo line was part of the original obligation under Term 11.⁴⁸ The negotiations dragged on for years with the Colonial Secretary, Lord Carnarvon, acting as an arbitrator and trying to help the parties identify a mutually acceptable compromise.⁴⁹

The two governments finally reached agreement in 1883, two years *after* the original deadline for construction of the CPR. By this time, the main line of the railway was nearing completion. The Settlement Agreement of 1883 stated that it represented a settlement of all claims by the province in respect of “delays in the commencement and construction of the Canadian Pacific Railway and in respect of the non-construction of the Esquimalt and Nanaimo Railway.” As for the island railway line,⁵⁰ British Columbia was to transfer a large portion of land on Vancouver Island to Canada which in turn agreed to grant this land to a private company for the construction of the line. The land would be exempt from taxation by the province as long as it was used for railway purposes. The federal government agreed to provide

a subsidy of \$750,000 for the construction of the Island railway. Under the terms of the Agreement, construction was to “commence forthwith” and was to be completed by June of 1887.⁵¹

The Agreement was ratified by legislation passed by the B.C. Legislature and by the Parliament of Canada,⁵² and the federal government contracted with a private company to construct the line. The terms of this second agreement were attached as a “schedule” to the federal legislation implementing the Settlement Agreement with British Columbia. Under this agreement, the contractors (known as the Dunsmuir Syndicate) were obliged to “construct, complete, equip, maintain and work continuously” the line of railway between Esquimalt and Nanaimo (section 3 of the Agreement). The contractors also agreed to maintain the railway “in good and efficient working and running order” (section 9) and to equip the railway in accordance with specifications set down by the federal government (section 10). The construction proceeded largely according to schedule, and in 1886 the 70-mile stretch of railway between Esquimalt and Nanaimo was completed and opened for traffic.

While the governments were reaching a settlement of these issues, the construction of the CPR main line was nearing completion. Under the terms of a contract signed in 1881, the federal government granted the Canadian Pacific Railway Company a subsidy of \$25 million and a land grant of 25 million acres to complete the railway by May of 1891. The company was also granted a permanent exemption from taxation by all levels of government.⁵³ In return, the company agreed to construct the line and to “thereafter and forever efficiently maintain, work and run the Canadian Pacific Railway.”⁵⁴ The terms of the contract with the CPR were approved and ratified by federal legislation.⁵⁵ In November of 1885, Donald Smith hammered in the last spike of the Canadian Pacific Railway and the project of a transcontinental railway was finally completed.

All the various agreements, contracts and statutes entered into regarding the completion of the Canadian Pacific Railway were incorporated into federal or provincial legislation and, accordingly, were legally binding. However, these statutes have been a source of considerable litigation over the years, primarily in relation to attempts by provinces to modify or cut down the tax exemptions which they contemplated.⁵⁶ But the important point, from a constitutional standpoint, is that none of these agreements or

statutes altered section 11 of the B.C. Terms of Union. Thus, while they may have been legally enforceable, they were not constitutionally entrenched. They could be amended or modified by Parliament or by a province, subject to the paramountcy of federal laws over provincial laws.⁵⁷

For example, in 1950 the Privy Council decided that the British Columbia legislation granting a tax exemption for railway lands on Vancouver Island could be amended or repealed by the provincial legislature.⁵⁸ Similarly, the federal legislation establishing the Canadian Pacific Railway was ordinary federal law which could be amended or repealed as the Parliament of Canada saw fit.⁵⁹ In this sense, the agreements did not themselves constitute a constitutional limitation on the jurisdiction of Parliament. The only constitutionally mandated obligation remained that specified in Term 11 of the B.C. Terms of Union: to secure the construction of the CPR.

While fulfilment of Term 11 proved a source of great controversy and dispute, there were no significant problems in satisfying the other transportation obligations in the B.C. Terms of Union. To fulfil its obligation to provide efficient mail service between Victoria and both San Francisco and Olympia, the federal government contracted with private carriers and paid any necessary subsidies. There is no record of any complaint from the province regarding the quality or level of the service. In 1925, the two governments agreed that further subsidy of the Victoria to San Francisco service was unnecessary. The federal subsidy of \$3,000 which had been paid towards this service was used by the province to improve mail service within British Columbia rather than for the San Francisco service.⁶⁰

While the province had agreed that the federal government should be “relieved of its obligation to maintain a subsidized steamship service between Victoria and San Francisco,”⁶¹ no change was made in the Terms of Union themselves. Accordingly, Term 4 of the Terms of Union continues to refer to the obligation to provide mail service on the Victoria–San Francisco route, even though this service was discontinued in 1925.

C. THE PRINCE EDWARD ISLAND FERRY

In the years following Confederation, the operation of the ferry service connecting P.E.I. with the Canadian mainland proved to be a source of considerable dispute. The federal government was obliged to provide this

ferry service under the Terms of Union with the province. The federal government attempted to fulfil this obligation by contracting with private companies to provide a subsidized service during the summer months. There appears to have been considerable dissatisfaction with the winter service which was provided by the government itself, using ferries operated by the Department of Marine and Fisheries.

In April of 1901, the province presented a memorial to the federal government alleging a failure by the Government of Canada to fulfil its obligation to provide a "continuous communication" between the Island and the mainland. The memorial claimed that the federal government's "solemn undertaking was systematically and continuously broken from the year 1873 to 1888 when for the first time in that latter year an adequate vessel was constructed and placed in service during the winter season."⁶² The province claimed damages in the amount of \$5 million for the alleged breach of the Terms of Union and asked that the claim be referred to a board of arbitrators. The federal government referred the claim to a committee which found that the federal government had failed to satisfy its obligations during the winter months of 1873 to 1887. The committee recommended that the province be paid an allowance of \$30,000 annually as compensation. The government accepted the recommendation and passed legislation authorizing the payments, stating that they would "be paid and accepted in full settlement of all claims of the said province against the Dominion of Canada on account of the alleged non-fulfilment of the terms of Union."⁶³

Following the 1901 settlement, there were further provincial complaints regarding the ferry service. In 1912, the Province presented another memorial seeking an increase in the annual subsidy set out in the 1901 legislation. The provincial claims were again referred to a committee for consideration. Following a series of negotiations, Parliament passed legislation authorizing an increase of \$20,000.⁶⁴ The federal government has continued to pay these subsidies annually in accordance with the terms of the legislation.

Following these two settlements, provincial complaints regarding the operation of the ferry service appear to have diminished. The federal government contracted with Northumberland Ferries Limited to operate a ferry service between P.E.I. and Nova Scotia and, beginning in 1923, employed CN Railway to operate the service between P.E.I. and New Brunswick. Until

the nation-wide strike by CN Rail in 1973, which resulted in a work stoppage of 10 days and a shutdown of the ferry service to New Brunswick, the federal government appears to have satisfied its obligations under the Terms of Union.

D. SUMMARY: POLITICAL OVER LEGAL ENFORCEMENT

The various transportation obligations reviewed here were all established by imperial statute or Order in Council. As such, they were legally binding obligations that could not be unilaterally altered by the federal government. While the obligations were binding, the striking feature of these provisions is that there was no legal mechanism for enforcing them. The only machinery contemplated by the *British North America Act* for settling disputes between Canada and the provinces in 1867 was section 142, relating to disputes over the division of debts and assets — a mechanism that did not apply to transportation obligations. In the absence of legal machinery for settling disputes, the parties would have to rely on negotiation, political pressure or the intervention of third parties to ensure that the obligations were carried out.

The early history of these obligations confirms that this was in fact the prevailing understanding of how they were to be enforced. In each instance when a dispute arose, the parties commenced political negotiations designed to reach a compromise acceptable to both sides. There was no attempt to invoke judicial involvement in the settlement of the dispute. Thus the dispute between British Columbia and Canada over the construction of the CPR was resolved through direct political negotiations as well as through the intervention of the imperial authorities. Similarly, when the Prince Edward Island government became unhappy with the manner in which the Island ferry service was being operated, it presented a brief to the federal government rather than to the courts.

A second observation is that, because the enforcement mechanism was political rather than legal, the political authorities of the time resolved their disputes without regard to the precise legal limits of the particular obligation. Under the B.C. Terms of Union, for example, the federal government was obliged to construct a railway to the “Seaboard” of British Columbia; it later agreed to finance construction of a railway line on Vancouver Island, even though such a line was not expressly required.

Similarly, the only constitutional obligation of the British Columbia government under Term 11 was to transfer land that was directly along the main line of railway in the province; it later agreed to transfer an additional block of some 3.5 million acres in the interior of the province when the land originally to be transferred proved to be of lower value than expected.

The same pattern is repeated in the case of the Intercolonial Railway and the P.E.I. Terms of Union. In the case of the Intercolonial, the constitutional obligation of the Canadian government was limited to the actual construction of the railway. Subsequently, however, the government adopted a policy of discounting freight rates on the line, even though there was no constitutional obligation to do so. In P.E.I.'s case, the federal government agreed to provide an annual subsidy of \$50,000 in perpetuity following provincial complaints over the quality of the ferry service.

In each case, there was no attempt to insist on the strict letter of the law; the overriding concern was to ensure that political compromises acceptable to all parties were achieved. Such reliance upon political mechanisms as an instrument of enforcing constitutional obligations was regarded at the time as entirely appropriate and straightforward. The political leaders of the 19th century would no doubt have been surprised by the modern tendency to rely on legal mechanisms to insist on strict adherence to constitutional commitments.

IV. COMPLETING CONFEDERATION 1949: TRANSPORTATION UNDERTAKINGS IN THE NEWFOUNDLAND TERMS OF UNION

The Newfoundland Terms of Union contain the most detailed set of constitutional obligations of any province in Canada. This is hardly surprising since the drafters of the terms had before them the precedents established by the Terms of Union with the other provinces as well as the interpretation of those terms.

The colony of Newfoundland at the time of Confederation in 1949 was particularly concerned about the status of the Newfoundland Railway which it had started to build in the late 19th century. The cost of its construction imposed a crippling financial burden on the island economy,

and the operation of the railway generated large financial losses.⁶⁵ Although the railway was taken over by the Newfoundland government in 1923, it remained unprofitable.⁶⁶ In the negotiations over the Terms of Union, the Newfoundlanders proposed that the federal government be responsible for any and all financial losses associated with the operation of the railway. The Canadian authorities' agreement is reflected in Term 31 of the Terms of Union:

31. At the date of Union, or as soon thereafter as practicable, Canada will take over the following services and will as from the date of Union relieve the Province of Newfoundland of the public costs incurred in respect of each service taken over, namely,

(a) the Newfoundland Railway, including steamship and other marine services;

During the negotiations, the Newfoundland authorities expressed some concern that the Terms of Union did not require the federal government to continue operating the Newfoundland Railway. The only obligation stated in Term 31 was to "take over" the railway and to pay for any losses which might arise from its operation. This suggests that the federal government would be in compliance with Term 31 as long as it ensured that the province did not have to assume any of the losses of the Newfoundland Railway. But the federal government apparently reserved the right to determine the level of service on the railway or, indeed, to shut it down entirely.

During the negotiations leading to Newfoundland's entry into Confederation, the Newfoundland authorities sought clarification of what was entailed by Canada "taking over" the Newfoundland Railway. Prime Minister St. Laurent replied:

During the course of our negotiations covering the final terms and arrangements for the union of Newfoundland with Canada a number of questions concerning Government policy were raised by your delegation and answered by the Canadian Government. In addition a number of temporary administrative arrangements were settled in order to facilitate the union.

It would not seem fitting to include in formal terms of union matters of this kind, since they are scarcely of a constitutional nature. I am therefore sending you the enclosed memorandum covering these

various items. While these will not form part of the Terms of Union, they contain statements of the policy and intentions of this Government if union is made effective by the approval of the Parliament of Canada and the Government of Newfoundland and confirmed by the Parliament of the United Kingdom.

Yours sincerely
Louis S. St. Laurent
[enclosure]

STATEMENTS ON QUESTIONS RAISED BY THE NEWFOUNDLAND DELEGATION

(xiv) NEWFOUNDLAND RAILWAY

After the date of Union, the Canadian National Railways will be entrusted with the responsibility of operating the Newfoundland Railway and Coastal Steamship Services, and it will be their responsibility to see that the services are furnished commensurate with the traffic offering.⁶⁷

The Sullivan Commission, in its report to the Newfoundland House of Assembly in 1978, argued that the letter from Prime Minister St. Laurent had the effect of modifying the Terms of Union. The Commission took the position that the letter obligated the federal government to maintain the Newfoundland Railway, regardless of cost, as long as there was reasonable demand for its services.⁶⁸

This interpretation of the effect of the letter seems rather doubtful. In the first place, while the courts will no doubt consider material such as parliamentary debates, government reports and other documents,⁶⁹ they tend to accord such material minimal weight in assigning meaning to the Constitution.⁷⁰ Secondly, St. Laurent's letter specifically distinguished between "government policy" and matters of a "constitutional nature." St. Laurent indicated that the policy and intentions of the government regarding the Newfoundland Railway "are scarcely of a constitutional nature . . . [and] will not form part of the Terms of Union." Thus St. Laurent himself suggested that the undertaking to operate the railway was a political rather than a constitutional commitment and that a conscious decision

was made *not* to include this duty in the Terms of Union. Thirdly, St. Laurent indicated that the policy of the government was to entrust the operation of the Newfoundland Railway to CN Rail and that “it will be *their* responsibility” [our emphasis] to see that the services were operated in accordance with the traffic offering. In effect, St. Laurent did not commit the government to ensure the continued or perpetual operation of the railway. The responsibility was to be imposed on CN by the government. The implication is that a subsequent government could decide to modify the terms and conditions under which CN would operate the railway.

This is precisely the approach taken in the federal Order in Council entrusting the operation of the railway to CN on condition that “such management and operation shall continue during the pleasure of the Governor in Council and be subject to termination or variation from time to time in whole or in part by the Governor in Council.”⁷¹ Under the terms of the Order, the federal government specifically contemplated the possibility that the operation of the railway would be “terminated . . . in whole or in part.” This provision can be contrasted to the terms set out with respect to the operation of the Canadian Pacific Railway, which specifically required the corporation to operate the railway in perpetuity. The absence of any such requirement in the 1949 Order In Council indicates that the government of the day did not believe that it had any constitutional obligation to ensure the perpetual operation of the railway.

A further significant consideration is that in Term 32(1) of the Newfoundland Terms of Union, the federal government specifically undertakes an obligation to maintain a specific transportation service, namely, the ferry service between North Sydney and Port aux Basques.⁷² This makes the failure to include any such statement with respect to the Newfoundland Railway all the more decisive.

Although this provision is similar to the ferry service provisions in the P.E.I. Terms of Union, it differs in its stated obligation to maintain the service “in accordance with the traffic offering.” The Sullivan Commission in 1978 speculated that there might be circumstances in which traffic would cease to “offer” and that, in such a case, the obligation to provide the service would cease. However, the Commission emphasized that such a situation would be exceptional and would arise only when there was no longer any demand for the service.⁷³ It is evident that the obligation to

provide the ferry service between North Sydney and Port aux Basques is very nearly absolute, similar in practical effect to the obligation regarding ferry service in P.E.I.

Between 1867 and 1873, railway rate regulation by the government was unknown. Rates were regarded as a matter of contract to be left to the determination of the market. Thus the various transportation obligations assumed by the federal government during this early period made no reference to them. By the 1940s, of course, the government had assumed a significant role in regulating railway rates. Thus it is hardly surprising that the Newfoundland Terms of Union would make reference to rates to be charged on the named transportation services. The relevant provisions in Term 32 were as follows:

(2) For the purpose of railway rate regulation the Island of Newfoundland will be included in the Maritime region of Canada, and through-traffic moving between North Sydney and Port aux Basques will be treated as all-rail traffic.

(3) All legislation of the Parliament of Canada providing for special rates on traffic moving within, into, or out of, the Maritime region will, as far as appropriate, be made applicable to the Island of Newfoundland.

The effect of Term 32(2) was to impose a constitutional cap on railway rates in Newfoundland and on the Port aux Basques ferry. Rates on the Island of Newfoundland itself were to be fixed in accordance with comparable rates in other parts of the Maritimes. Thus, even though transportation in Newfoundland might be more difficult or costly, rates were not permitted to move above the level of rates throughout the Maritimes.⁷⁴ Further, Term 32(2) stated that the rates for rail traffic moving on the Port aux Basques ferry were to be set as if the traffic were moving on land rather than by ship.

Term 32(3) has a slightly different impact. This term did not impose an absolute rate cap. Rather, it simply required that any legislation enacted by the Parliament of Canada providing for special or preferential rates for the Maritime region would also be applied to Newfoundland. There is no obligation to enact such legislation, however. Nor is the federal government constitutionally barred from repealing any legislation which it chooses to

enact, as long as it treats Newfoundland on a footing identical to that of the other Maritime provinces. Of course, given the long history of preferential freight rates for the Maritimes, dating back to the operation of the Inter-colonial Railway, the possibility of repealing such legislation is perhaps more theoretical than real. The point is simply that nothing in the Terms of Union prevents Parliament from amending its legislation, as long as any preferential rates applying in the Maritimes also apply in Newfoundland.

V. THE PRESENT STATUS OF THE UNDERTAKINGS: THE MOVEMENT FROM POLITICAL TO LEGAL ENFORCEMENT

As we have seen, the original method for enforcement of federal government transportation obligations was political rather than strictly legal. A province that was unhappy with the way in which the federal government was carrying out its responsibilities would typically complain to the federal or imperial authorities. Eventually some compromise solution would be proposed which met the concerns of both levels of government.

In recent years, there has been a slow but discernible shift in the approach to these various constitutional entitlements. Although the transportation obligations set out in the Canadian Constitution continue to be the subject of political negotiations and bargaining compromise between the two levels of government, there is an increasing tendency to seek judicial and legal enforcement of these obligations, rather than to rely exclusively on political avenues of redress. The courts have demonstrated a willingness to take on responsibility for enforcing these obligations, and have assumed an increasingly significant role in their interpretation.

A watershed in this regard was litigation undertaken by the Prince Edward Island government in the mid-1970s following the shutdown of part of the Island ferry during a labour dispute.⁷⁵ This case established the proposition that provincial governments could seek enforcement of transportation obligations through the courts. The case also attempted to define the precise legal nature of the obligations and the extent to which it was open to private citizens to seek legal enforcement.

A. THE PRINCE EDWARD ISLAND FERRY

A nation-wide legal strike by employees of CN Rail in August of 1973 resulted in the shutdown of the ferry service between New Brunswick and P.E.I. for 10 days. Although ferry services were still operating between Nova Scotia and P.E.I. as well as air service to and from the Island, since it was the height of the tourist season, these alternative services were unable to meet the demand. As a result the shutdown had a major negative impact on the Island economy.⁷⁶

The province commenced an action in the Federal Court seeking damages for the losses suffered by the Island during the strike. The province alleged that the federal government had breached its obligation under the P.E.I. Terms of Union to maintain an efficient ferry service linking the Island and the mainland and that the province had a right to compensation for the resulting losses.

The trial judge, Mr. Justice Cattanach, agreed with the province that the federal government had breached its constitutional obligation to provide an efficient ferry service. Holding that an “efficient” service is one that is reasonably capable of meeting the demand for the service, he concluded that there had been a breach of this undertaking during the strike since the service was “wholly inadequate for the need at that time.”⁷⁷

However, Mr. Justice Cattanach stated that this breach of obligation did not give rise to an action for damages by the province. In arriving at this conclusion, Cattanach J. distinguished between actions for declaratory relief and actions for damages. He was prepared to grant a mere declaration setting forth the rights and obligations of the Dominion vis-à-vis the province,⁷⁸ but he concluded that the province could not succeed in an action claiming monetary compensation for breach of the obligation to provide ferry service. His reasoning was that the obligation was one owed to the public generally rather than for the benefit of any particular individual or class of individuals. Because the obligation had been created for the general public good, Cattanach J. reasoned that no particular person or even province had a right to seek damages for a breach of the obligation.⁷⁹

The Federal Court of Appeal agreed with Justice Cattanach that the federal government had breached its constitutional obligation to provide an efficient ferry service. However, a majority of the Court disagreed with his

conclusion that the province could not bring an action for damages. Chief Justice Jaccett concluded that the effect of the Terms of Union was to impose a legal duty on "Canada" in favour of "the Province" of Prince Edward Island. He reasoned that when there is a statutory right to have something done, there is an implied right to be compensated for a breach of such right.⁸⁰ Jaccett C. J. concluded that the province of Prince Edward Island, as opposed to the residents of the Island themselves,⁸¹ had a right to be compensated for losses sustained due to the interruption of ferry service.

In a separate opinion, Mr. Justice Le Dain agreed with this conclusion. He stated that breach of the constitutional obligation to provide ferry service did not permit individual citizens to sue for monetary compensation. However, he was of the view that the provincial government could sue for any loss directly caused to it by the failure to provide efficient ferry service.⁸²

In summary, this case established the following three propositions:

1. The constitutional transport obligations set out in the P.E.I. Terms of Union are legally enforceable;
2. In the event that the obligations are not fulfilled, the provincial government has a right to monetary compensation from the Government of Canada for losses resulting from the breach; and
3. It is unlikely that individual citizens have any right to compensation for losses which they might have suffered as a result of the breach of the obligation; the right to compensation is apparently limited to the provincial government alone.

It is important to consider whether the principles set down in *Canada v. P.E.I.* have been modified in any way by the enactment of the *Constitution Act, 1982*. Section 52(1) of the *Constitution Act, 1982* provides that "The Constitution of Canada is the supreme law of Canada, and any law that is inconsistent with the provisions of the Constitution is, to the extent of the inconsistency, of no force or effect." The "Constitution of Canada" is defined in section 52(2) and it includes the BNA Act, 1867, the *British Columbia Terms of Union*, the *Prince Edward Island Terms of Union* and the *Newfoundland Act*.⁸³ It is evident, in other words, that all of the various provisions establishing transportation obligations of the federal government are included within the meaning of the term the "Constitution of Canada."

Thus, under the terms of section 52, any law that is inconsistent with the constitutional provisions establishing these transportation obligations is of “no force and effect.” Any attempt by the Parliament of Canada to modify or reduce its constitutional obligations to provide transportation services is legally invalid. It is clear that a provincial government could seek a declaration of invalidity in accordance with section 52 of the *Constitution Act, 1982*.⁸⁴ The only remaining question is whether private individuals could also bring such an action. The Federal Court of Appeal in the *Canada v. P.E.I.* case was of the view that private individuals could not. But an action under section 52 of the *Constitution Act, 1982* is not an attempt to obtain monetary compensation; section 52 simply contemplates the court declaring invalid any law that is inconsistent with the Constitution of Canada.

Courts have taken an increasingly liberal attitude to the question of who has a right to sue for a declaration of constitutional invalidity. The present rule is that a private citizen can maintain an action for a declaration that legislation is invalid if that person can show “that he is affected by it directly or that he has a genuine interest as a citizen in the validity of the legislation and that there is no other reasonable and effective manner in which the issue may be brought before the Court.”⁸⁵

There seems to be no reason why this general rule should not be applied in the case of the constitutional provisions relating to transportation. Under this approach, the class of persons who could seek a declaration relating to a particular transportation obligation would be extremely broad and open-ended. Any citizen who was a user or even a potential user of a particular transportation service is affected by a decision to reduce or eliminate that service. It would appear, therefore, that even potential users of the constitutionally mandated transportation services would have legal standing to seek a declaration relating to that service. Nor would the class of potential users (and thus litigants) be limited to the residents of a single province.

As the Federal Court indicated in the *Canada v. P.E.I.* case, the obligation to provide ferry service to P.E.I. is for the benefit of residents on the mainland as well as those on the island.⁸⁶ The object of guaranteeing the service is to ensure effective communication *between* the residents of the various provinces. The same can be said of the other transportation undertakings that link provinces. It would appear that any citizen who was a user or

potential user of a constitutionally protected transportation service could challenge a decision to reduce that service below the level guaranteed by the Constitution.

The implications of this conclusion may not be as dramatic as might at first be assumed. First, nothing in section 52 of the *Constitution Act, 1982* grants citizens the right to seek monetary compensation in the case of constitutional invalidity. Thus the conclusion of the Federal Court of Appeal in the *Canada v. P.E.I.* case that individual citizens cannot seek financial compensation for breach of a transportation obligation has not been affected by the *Constitution Act, 1982*.

Secondly, and more importantly, the question of standing to sue is a subsidiary one: the determinative question is the scope of the legal obligations themselves. If, as argued earlier, the transportation obligations of the federal government are relatively narrow and circumscribed, then the fact that a broad class of citizens can seek to enforce those obligations is of secondary importance. The key issue, in other words, is the extent of the obligation, rather than who can enforce it.

B. THE PRINCE EDWARD ISLAND RAILWAY

In 1989, the National Transportation Agency granted an application from CN to abandon all the rail lines on Prince Edward Island.⁸⁷ The Prince Edward Island government challenged this Order on the grounds that the P.E.I. Terms of Union imposed an obligation on Canada to operate the railway. The P.E.I. argument relied on the wording of the federal obligation to provide a ferry service: ferry service was to be maintained “thus placing the Island in continuous communication with the Intercolonial Railway and the railway system of the Dominion.” The province claimed that this implied an obligation to operate the Island railway as well as the ferry service; without a railway, there would be no way of connecting with the railway system of the Dominion. The Federal Court of Appeal unanimously rejected this argument.⁸⁸ Chief Justice Iacobucci (as he then was) held that the only obligation was to provide a ferry service. The reference to the “railway system of the Dominion” in the Terms of Union did not extend this obligation but merely described its effect. Thus there was no constitutional bar to a decision to shut down the railway.⁸⁹ It is clear from this decision that the only ongoing transportation obligation relating to Prince Edward

Island is to maintain an efficient ferry service, in accordance with the principles outlined in *Canada v. P.E.I.*⁹⁰

C. THE NEWFOUNDLAND TRANSPORTATION OBLIGATIONS

At the time of Newfoundland's entry into Confederation, the Newfoundland Railway was the only reliable means of overland transportation across the province. On a narrow gauge rail with many steep grades and severe curves, it normally took between 22 and 30 hours to make the 547-mile trip from St. John's to Port aux Basques. Passenger traffic on the railway began to fall off in 1960 and then dropped off substantially after 1965 when the Trans-Canada Highway was completed. The cost of converting the railway to standard gauge or otherwise to improve it was found prohibitive.⁹¹

In 1967, CN applied to the Canadian Transport Commission (CTC) to discontinue the passenger train service and begin operating a bus service instead. The Roadcruiser bus service was said by CN to be more efficient and convenient than the train service it was to replace. The Railway Transport Committee of the CTC approved the application but adopted what it termed a "large and liberal interpretation" of the Newfoundland Terms of Union so as to give it "the flexibility that changing or unforeseen circumstances may require."⁹² According to the Committee, there was a "presumption" under the Terms of Union that public transportation service for passengers between St. John's and Port aux Basques "should be assured so long as it is required by public convenience and necessity." Because of this presumption, the Committee attached a number of conditions to its Order, the most important being that the CN must maintain the bus service "as long as a requirement for passenger service continues."⁹³

This interpretation of the transportation obligations of the federal government was indeed both expansive and novel. The relevant provisions in the Terms of Union merely specify that the federal government is to take over the Newfoundland Railway. There is no stated requirement to operate the railway, in contrast to the provision requiring continued operation of a ferry service across the Cabot Strait. The Railway Committee did not interpret the Terms of Union as requiring the maintenance of a passenger rail service *per se*. Rather, it held that the federal government was required to maintain some form of public transportation service which would meet the needs of the residents of Newfoundland.

Canadian National has continued to operate the Roadcruiser bus service in accordance with requirements established by the Canadian Transport Commission and its successor, the National Transportation Agency. While a wholly intra-provincial bus service, it falls under federal jurisdiction due to its close integration with the Newfoundland Railway system.⁹⁴

Although CN discontinued its passenger rail service in Newfoundland in 1969, it continued its freight service on the Newfoundland Railway system. But freight service became increasingly unprofitable, particularly as a result of growing competition from truck transportation on the Trans-Canada Highway. A Provincial Commission of Inquiry into Newfoundland Transportation appointed in the mid-1970s found that the railway could not continue as a viable service.⁹⁵ The Commission recommended that the railway system be phased out entirely over a 10-year period. Both the provincial and federal governments rejected this recommendation and instead undertook a major revitalization freight transport program for the railway. Despite these efforts, the railway's losses mounted to over \$40 million annually in the early 1980s while its market share continued to decline. By 1987, the railway's share of total freight traffic in the province had fallen to 20 percent, and was projected to decline further.⁹⁶

In 1988, the federal and provincial governments signed a Memorandum of Understanding providing for the phase-out of the Newfoundland Railway on September 1, 1988. Under the terms of the Agreement, the federal government was to provide over \$800 million for the improvement of roads and port facilities as well as for labour and community adjustment.⁹⁷ The Memorandum of Understanding also provided that the federal payments were offered "in full satisfaction of all Canada's constitutional obligations related to railways on the Island of Newfoundland. . . ." [paragraph 10(1)]. Under the Agreement, the province stated that Canada had met all of its constitutional obligations relating to railways on the Island of Newfoundland. In accordance with the terms, the Newfoundland Railway ceased operations, and has been largely dismantled; track and ties have been lifted and the right-of-way returned to its original state.⁹⁸

It is a basic principle of Canadian constitutional law that governments cannot alter the Constitution through mere agreement.⁹⁹ Indeed, the Supreme Court of Canada has recently determined that intergovernmental agreements are subject to repeal or abrogation by statute.¹⁰⁰ Thus it is clear that the

Memorandum of Understanding cannot have the effect of altering any constitutional obligations which the federal government might have with respect to the operation of the Newfoundland Railway. However, as this report suggests throughout, the federal government has never been under a constitutional obligation to maintain or operate the Newfoundland Railway. This interpretation flows from the Terms of Union themselves, which merely provide that the Government of Canada is to “take over” the railway. The absence of any requirement to maintain the service means that there is no constitutional objection to a decision to close down the railway.

This interpretation of the Terms of Union was explicitly endorsed by the Premier of Newfoundland, the Honourable Brian Peckford, at the time of the signing of the Memorandum of Understanding. Premier Peckford offered the following comments:

We have assessed the legal intent and obligations imposed on the federal government by the Terms of Union. It is clear that the Government of Canada does not have a legal obligation to operate a railway in Newfoundland forever. We have always felt, however, that the federal government does have an obligation to ensure that there is a viable transportation system in this province. This agreement today constitutes our mutual recognition that this comprehensive transportation package meets that obligation.¹⁰¹

In the period since the shutdown of the Newfoundland Railway, the National Transportation Agency has had occasion to interpret and apply the transportation obligations in the Terms of Union. In its conclusions, the Agency has simply assumed that the 1988 shutdown was perfectly lawful and consistent with the Terms of Union. For example, the Agency recently was asked to interpret the meaning of section 32(2) of the Terms of Union in light of the shutdown of the railway. There was no suggestion that the decision to shut down the railway was in any sense a violation of the Terms of Union.

All of these factors confirm the conclusion that the federal government had no constitutional obligation to maintain the Newfoundland Railway. Accordingly, it would seem that the termination of the railway in 1988 did not violate any of the constitutional obligations of the federal government under the Newfoundland Terms of Union.

The remaining question is how the shutdown of the railway may have affected the other transportation obligations in the Terms of Union. Of particular interest in this regard is Term 32(2), which provides that “for the purpose of railway rate regulation the Island of Newfoundland will be included in the Maritime region of Canada. . . .” The problem is that, while Term 32(2) specifically refers to “railway rate regulation,” there is no longer a railway system in Newfoundland. Does that mean that the shutdown of the railway has somehow rendered Term 32(2) redundant or inapplicable? The National Transportation Agency’s unanimous decision was that Term 32(2) continued to apply despite the shutdown of the railway.¹⁰² Although there is no longer a railway in Newfoundland, there are still “railway rates” which are to be developed by CN and applied in accordance with Term 32(2). The Agency ruled that Terms of Union rates should be developed using rail mileage through North Sydney to Port aux Basques and onto St. John’s as if the Newfoundland Railway were still in place. The Agency made it clear, however, that such rates constitute a “ceiling” only and there is no prohibition on CN charging rates lower than the Terms of Union rate, as long as such rates satisfy the other requirements of the *National Transportation Act, 1987*.

While the Agency has determined that Term 32(2) continues to apply as a matter of strict law, other developments in the transportation marketplace are reducing its practical impact and importance. Term 32(2) only protects marine traffic moving over a named route from North Sydney to Port aux Basques. At the time of the negotiation of the Terms of Union in 1949 this was the main marine connection between Newfoundland and the Canadian mainland. However, in recent years there has been a dramatic shift in traffic patterns to the Island of Newfoundland such that only about 25 percent of CN’s traffic to Newfoundland now moves over the North Sydney–Port aux Basques gateway. The vast bulk of CN’s traffic moves between Halifax and St. John’s, a route that is more efficient and less costly for shippers. Technology and the transportation marketplace are thus rendering the protections of Term 32 increasingly redundant from a practical point of view.

A further development which has reduced the practical significance of Term 32(2) is the advent of confidential contracting under the *National Transportation Act, 1987*. The vast bulk of goods now moves at rates that are significantly less than published railway tariffs. Upwards of 65 percent of revenue traffic moving to Newfoundland now moves under confidential contracts.¹⁰³ Because the terms of these contracts are private, it is

becoming increasingly difficult to establish a benchmark Terms of Union rate that reflects the actual costs of moving goods to and from Newfoundland.¹⁰⁴ It would appear that the majority of goods moving to Newfoundland does so at rates well below the constitutional “ceiling” established by the Terms of Union.

The Newfoundland situation can be contrasted with that prevailing in Prince Edward Island. Under the P.E.I. Terms of Union there is no restriction on the rates that may be charged for the ferry service linking the Island with the mainland. Accordingly, any limitations on the rates for the P.E.I. ferry service are political and economic, rather than constitutional. Only Newfoundland has an explicit constitutional guarantee with respect to freight rates on the Sydney to Port aux Basques ferry and on the Island of Newfoundland itself.

D. THE VANCOUVER ISLAND RAILWAY

As previously described, in 1883 the governments of Canada and British Columbia agreed to settle their differences over the delays in the construction of the Canadian Pacific Railway. As part of this settlement, there was an agreement as to the manner of construction of a railway on Vancouver Island between Esquimalt and Nanaimo. The province agreed to make certain land grants in favour of the federal government on completion of the railway. The Government of Canada was to designate and contract with the persons who would build the railway, to transfer the land grant to these persons, and to contribute \$750,000 towards the cost of construction. The Canada-B.C. Agreement, which was ratified by legislation passed by both governments, made no provision for the manner in which the railway was to be operated upon its completion.

As already mentioned, pursuant to the Agreement, Canada contracted with a third-party syndicate (the Dunsmuir Syndicate) to build and operate the railway “continuously and in good faith.” The Canada-Dunsmuir Agreement was appended as a schedule to the federal Act of 1884 ratifying the settlement with the province.

The 70-mile stretch of railway was completed in 1886. The rights and obligations of the Dunsmuir Syndicate were transferred to the Esquimalt and Nanaimo Railway Company and later to the Canadian Pacific Railway

Company. Upon CPR assuming control in 1905, the railway was declared to be a work for the general advantage of Canada and came under federal regulatory authority.

Rail service continued to be offered until late 1989, when the federal Cabinet terminated passenger rail service on this line. The province sought a declaration from the British Columbia Supreme Court that the federal government was obliged to provide the service in perpetuity. The province argued that the obligation to operate the railway “continuously and in good faith” was constitutionally binding on the federal government. The province had to surmount two very large obstacles to succeed with this argument: the first was that nothing in the B.C. Terms of Union obliged the federal government to operate either the Canadian Pacific Railway or the branch line on Vancouver Island.¹⁰⁵ Secondly, nothing in the 1883 Settlement Agreement between Canada and British Columbia obliged the federal government to operate the Island railway. The only commitment regarding the operation of the Island railway is found in the Agreement between Canada and the Dunsmuir Syndicate. The promise to operate the railway “continuously and in good faith” was made to the federal government rather than to the province.

Despite these seemingly persuasive objections to the provincial argument, Mr. Justice Esson of the B.C. Supreme Court found that the federal government had a perpetual obligation to operate the Vancouver Island Railway. With respect to the 1883 Canada-B.C. Settlement Agreement, Esson J. acknowledged that the federal government had not promised to operate the railway continuously. However, he was of the view that the province had relied on the undertaking of continuous operation which the Dunsmuir Syndicate had given to the federal government. Therefore, the promise to operate the railway continuously was a benefit “which the Dominion impliedly offered to maintain for the benefit of the province.”¹⁰⁶ This “implied offer” became binding on the federal government when the province carried out its part of the bargain, particularly the “very onerous terms relating to the land grant.”

Mr. Justice Esson went on to conclude that the 1883 Agreement had “constitutional force” and could not be unilaterally amended by the federal government. The basis for this conclusion appears to be that the 1883 Agreement was “part of the constitutional compact under which British Columbia

became part and parcel of the Dominion.”¹⁰⁷ He further relies on the “intentions” of the two governments in 1871 and 1883 as a basis for finding an obligation to operate the railway indefinitely:

In 1871 no one would have thought that, if the railway was built, it would not operate in perpetuity. As between the two levels of government, it was unnecessary to stipulate that both would have the benefit of the covenant to operate — enough that the promise was given to the Dominion by the contractors and the railway company. I infer from all the circumstances that the province relied upon the Dominion to enforce that obligation.

What is most striking about this passage is that it makes no distinction between the construction of the island railway and the main line of the CPR. According to Esson J., the assumption *in 1871* was that if the railway was built it would operate in perpetuity. But the railway that was under contemplation in 1871 was not the Island railway; rather, it was the transcontinental railway linking B.C. with the rest of Canada. The logical implication flowing from Justice Esson’s reasoning is that the federal government has an obligation to operate *the whole of the Canadian Pacific Railway system in perpetuity*, not simply the 70-mile stretch of rail on Vancouver Island.

This conclusion is not as improbable as it might at first glance appear. It is important to remember that the CPR gave an undertaking of continuous operation to the federal government similar to that made by the Dunsmuir Syndicate. The federal legislation establishing the CPR required the corporation to run the whole of the transcontinental railway system in perpetuity; the wording in the case of the CPR is even more definitive and unambiguous than it was with respect to the Island railway.¹⁰⁸ Thus, if B.C. is entitled to rely on the undertaking with respect to continuous operation of the island railway, there would seem to be no reason preventing it from also relying upon the promise of perpetual operation of the CPR.

Thus Mr. Justice Esson’s interpretation of the B.C. Terms of Union and the 1883 Settlement Agreement has important implications for the whole of the CPR, not just the rail line on Vancouver Island. Under the approach adopted by Mr. Justice Esson, a variety of other federal government orders shutting down or reducing passenger service on the CPR would come into question.

Yet there are a number of difficulties associated with the reasoning of the Court in this case. These difficulties include the following:

- (i) The legislative history of the B.C. Terms of Union indicates that the federal government's obligation was to construct the CPR, but not to operate it; of particular significance is the fact that there are references to a requirement to "maintain" services elsewhere in the Terms of Union, but no such obligation was included with respect to the CPR;
- (ii) This interpretation places British Columbia on the same footing as the other provinces in Canada; the federal government has assumed responsibility for constructing or taking over a variety of other railways, but in no instance has it agreed to operate these railways in perpetuity. The decision to shut down the P.E.I. and Newfoundland railways has been accepted by the courts and/or the province concerned as perfectly lawful. To hold otherwise in the case of British Columbia is to place that province in a preferred and unique position with regard to railway operation; and
- (iii) Whatever may have been the intentions surrounding the Canada-B.C. Settlement in 1883, the Agreement did not amount to an amendment of the Terms of Union. Governments cannot alter the Constitution through mere agreement.¹⁰⁹ Thus, even assuming that the 1883 Canada-B.C. Agreement included a promise to operate the railway in perpetuity, such a promise is not constitutionally binding. The 1883 Agreement is not part of the "Constitution of Canada"¹¹⁰ and thus cannot amount to a fetter on the legislative authority of the Parliament of Canada.

Notwithstanding these objections, a five-member panel of the British Columbia Court of Appeal unanimously upheld Mr. Justice Esson's conclusions in a judgement handed down on October 4, 1991.¹¹¹ The Court of Appeal concluded that Canada had a continuing constitutional obligation to British Columbia to ensure the maintenance of passenger and freight rail service on the rail line between Victoria and Nanaimo. While the Court of Appeal found that the obligation was "continuous" it refused to describe the obligation as "perpetual";¹¹² it also indicated that the service could be discontinued with the agreement of the Government of British Columbia.

The major focus of the Court of Appeal's reasoning dealt with whether the construction of the Vancouver Island Railway was part of the original Term 11 obligation of Canada. The Court of Appeal reasoned that the construction

of the Island portion of the railway was indeed included within the constitutional obligation defined by Term 11. The Court of Appeal suggested that the wording of Term 11, which referred to the construction of a railway to the “Seaboard” of British Columbia, was ambiguous; in the Court’s view, this might be read so as to include the construction of a railway line on Vancouver Island. The Court of Appeal relied particularly on an 1873 Order in Council passed by the Government of Canada fixing Esquimalt as the western terminus of the Canadian Pacific Railway. In the Court’s view, this was the “best evidence of the intention and understanding of the parties about Canada’s Term 11 obligations.”¹¹³

Having found that the obligation defined by Term 11 included the construction of the rail line on Vancouver Island, the Court of Appeal went on to find that the 1883 Settlement arrangements were of a “constitutional nature.” In the Court’s view, Term 11 was a “skeletal” provision which was “worked out” through the various covenants associated with the 1883 Settlement.¹¹⁴ These covenants included the undertaking given by the Dunsmuir Syndicate to operate the Vancouver Island Railway “continuously.” This meant that this covenant was of a constitutional nature and could not be varied except with the consent of the Government of British Columbia.¹¹⁵

To what extent does the Court’s reasoning in this case have wider implications for other transportation obligations of the Canadian government? As already noted, the reasoning of Mr. Justice Esson would seem to imply an obligation to maintain perpetual operation of the whole of the Canadian Pacific Railway. This implication is strengthened by the reasoning of the Court of Appeal in this case. The Court of Appeal reiterates at numerous points in its judgement that the island railway is “constitutionally indistinguishable from the Mainland railway.”¹¹⁶ The Court of Appeal also implies that Term 11 itself carries with it some kind of continuing obligation:

Canada assumed a constitutional obligation to British Columbia, and indeed to all of Canada, to ensure the continuation of the arrangements made to carry the Terms of Union into effect for the benefit of all its citizens. *It is untenable, in our view, to argue that Term 11 was spent once the last spike on the Mainland Railway was driven.* [emphasis added]¹¹⁷

The Court's theory that Term 11 is a "skeletal" provision only which was to be "worked out" through subsequent enactments also supports the conclusion that the whole of the CPR must be operated in perpetuity. As noted above, the 1881 statute creating the CPR stated that the railway must be operated "forever" by the company. This statutory language could be said to be evidence of what the Court of Appeal identified as an obligation to "ensure the continuation of the arrangements made to carry the Terms of Union into effect for the benefit of all its citizens."

In summary, the implications of the litigation surrounding the Vancouver Island Railway extend far beyond the particular issues raised in the case. The reasoning of both the trial judge and the Court of Appeal supports the conclusion that the federal government has a constitutional obligation to maintain the whole of the Canadian Pacific Railway.¹¹⁸ Given these very broad implications, it would seem an appropriate matter for review by the Supreme Court of Canada.¹¹⁹

The recent litigation on this issue has injected some uncertainty into the precise scope and meaning of Term 11 of the B.C. Terms of Union. The other terms seem straightforward: under Term 4, the federal government is obligated to provide efficient mail service between Victoria and both San Francisco and Olympia. As we noted earlier, the federal approach was to contract with private operators to provide this service and to pay subsidies. Beginning in 1925, a federal subsidy directed towards supporting a ferry service to San Francisco was used by the province to support mail service within the province.

In the 1970s, British Columbia complained that the federal subsidies paid to support ferry service in British Columbia were lower than comparable subsidies paid in the Maritimes. As a result, in 1977 the province and the federal government entered into a new arrangement for ferry subsidies. Under the 1977 Subsidy Agreement,¹²⁰ the federal government agreed to pay a block grant of some \$8 million annually to the province for support of ferry service. In return for this subsidy, Canada was to be relieved of "any and all obligations for the provision of subsidy or other financial assistance over and above the subsidy provided for in this Agreement" (section 5). It would be up to the province to determine how the federal grant should be used. However, the province agreed to "assure reasonable and adequate service and appropriate supervision thereof" in B.C. coastal waters

(section 4(2)). The province was also obliged under the Agreement to “place appropriate passenger vessels in service . . . to give effective links where required on the coast between communities and principal water and air services” (section 4(3)).

An effective and efficient ferry system has been maintained in the province pursuant to this Agreement. Thus the requirements of the Terms of Union in this regard are currently being satisfied. But the constitutional obligation to provide the service remains that of the federal government rather than the province. The provisions in the Subsidy Agreement imposing obligations on the provincial government are subject to the constitutional requirements of the Terms of Union, which continue to apply.

VI. CONCLUSIONS

The first important conclusion is that the ongoing obligations of the federal government are relatively limited and circumscribed. All of the obligations relating to the construction of various railways in different provinces of Canada have long since been satisfied. As suggested, there is no continuing obligation to operate any of these rail services¹²¹ nor is there any constitutional limitation on the manner in which the services are to be provided. The only continuing obligations of the federal government would appear to fall into two categories:

- (i) The federal government is obliged to provide the ferry services guaranteed to British Columbia, Prince Edward Island and Newfoundland under the Terms of Union for these particular provinces. Further, it is obliged to maintain passenger and freight rail service on the Island railway on Vancouver Island, pending a final court determination of this issue;
- (ii) There is a limitation on the rates which can be charged to Newfoundland in accordance with terms 32(2) and 32(3) of the Newfoundland Terms of Union. It should be noted that this is the only constitutional limitation respecting rates; there is no constitutional constraint on cost recovery in respect of the other mandated ferry services.

The relevant province is entitled to undertake legal action to enforce these ongoing obligations. In addition, the province can seek monetary compensation for any losses which it might have suffered due to interruption of the

guaranteed service. It would appear that private citizens who are affected by breach of a constitutional obligation are also entitled to bring legal action to enforce it. However, it is unlikely that a private citizen would be permitted to obtain monetary compensation for any losses suffered due to a breach of a constitutional obligation. The private citizen could only obtain a declaration of his or her rights with respect to the transportation service in question.

If this first conclusion is correct, then these limited transportation obligations in the Canadian Constitution ought not to be a significant concern for transportation policy makers. Their extremely restricted scope and impact clearly do not represent a significant impediment to the development of a modern and integrated transportation system for Canada.

In one sense, it might be concluded that the inclusion of these transportation obligations in the Constitution was a success; the existence of these undertakings played a key role in the creation of the Canadian State. The critical feature of the undertakings was their binding and enduring character. Because they were incorporated in imperial legislation, successive Canadian governments were obliged to make good on the original undertaking. This provided the colonial leadership in the 1860s and 1870s with the assurances it needed to support political union.

But there are disadvantages to including provisions of this type in a constitution. A constitution is intended to provide a general framework which can be adapted to fit the changing needs and circumstances of state and society. The transportation undertakings set out in the Canadian Constitution are not of this general character. Rather, they set out quite precise commitments to provide certain transportation services, including in some cases the time and manner in which the service is to be carried out. The problem with this type of constitutional provision is not simply that it is out of character with the generality of the constitution as a whole. The real difficulty is that the more specific a constitutional provision, the more difficult it is to adapt that provision to changing circumstances or needs.

One risk is that the provision will simply be rendered meaningless or superfluous by the changing current of events. An illustration of this is the guarantee of preferential rates on the North Sydney to Port aux Basques ferry in

the Newfoundland Terms of Union. While this route was once the main gateway to Newfoundland, changes in technology are rendering it increasingly marginal to Newfoundland's transportation system.

But a second risk, by far more worrisome, is that the courts will attempt to apply a particular provision to changing circumstances in a novel or unforeseen way. One possibility is that the courts will insist that a particular mode of transportation must be used, regardless of its relative cost or efficiency. In effect, the courts would attempt to "freeze" the evolution of the transportation system and prevent the movement to more efficient and cost-effective modes. For example, there has been discussion over the years of the possibility of establishing some form of fixed link between Prince Edward Island and the Canadian mainland. Were this fixed link ever constructed, it might render the continuation of ferry service over the same route superfluous. However, the existence of a constitutionally mandated obligation to provide a ferry service would likely prevent governments from discontinuing the service, even though it would have outlived its usefulness.

Of course, even in a case where the courts required the continued use of an inefficient or uneconomic mode of transportation, it would still be open to governments to amend the Constitution.¹²² On the other hand, constitutional amendment in Canada is never a simple or straightforward matter, as the events of the past decade have shown.

On balance, therefore, there would appear to be significant disadvantages associated with the constitutional entrenchment of entitlements to specific transportation services. It is no accident that the practice of including such obligations in the Canadian Constitution has been limited to a means of securing the entry of individual provinces or groups of provinces to Canada. Once a province was created, however, there is no instance where its constitutional entitlement to transportation services has been expanded. Instead, transportation services involving particular provinces have been provided for through ordinary legislation or federal-provincial negotiation and agreement. There has been no willingness to entrench any further entitlements in the Constitution. The inclusion of these provisions is clearly anomalous, for the reasons given above. The exceptional character of the existing constitutional transportation obligations is all the more apparent in light of the enactment of the *Constitution Act, 1982*. The entitlements to specific

transportation obligations are now included as part of the "Supreme law of Canada" and any laws which are inconsistent with this fundamental law are of no force and effect.

It is to be expected that provincial entitlement to transportation services will continue to be dealt with largely through ordinary legislation and federal-provincial agreement. No province has sought to entrench further entitlements to transportation services in the Constitution. At the same time, the recent Supreme Court case dealing with the enforceability of federal-provincial agreements¹²³ may cause some provinces to demand some mechanism to bind the federal government to meet its contractual obligations.

The answer to this concern is not, as has sometimes been suggested, to entrench federal-provincial agreements directly in the Constitution. As argued above, the constitutional entrenchment of what amount to contractual commitments between the two levels of government can have effects that are both unanticipated and unsatisfactory. A better solution to this potential problem¹²⁴ is to use the mechanism which was provided under the Meech Lake Accord with respect to immigration: any province which negotiated an agreement with the federal government regarding immigration could have the agreement protected from unilateral amendment by the federal government.¹²⁵ This approach, which falls short of entrenching the agreement itself as part of the Constitution, could be applied more generally. In transportation, for example, it could provide a mechanism to ensure that federal government undertakings were legally enforceable. It would also provide an answer to any provincial concerns regarding the uncertain legal status of agreements with the Government of Canada.

ENDNOTES

I am indebted to Mr. Christopher Morrison, a member of the Osgoode Hall class of 1993, for his excellent research assistance in the preparation of this paper.

1. G. P. de T. Glazebrook, *A History of Transportation in Canada*, Vol. 1, (Toronto: McClelland and Stewart, 1964), p. xiii.
2. See, for example, the guarantee in the B.C. Terms of Union that Canada will provide "an efficient mail service, fortnightly, by steam communication between Victoria and San Francisco, and twice a week between Victoria and Olympia; the vessels to be adapted for the conveyance of freight and passengers." B.C. Terms of Union, section 4.

3. The single most significant and costly transportation obligation was, of course, the undertaking to construct a transcontinental railway linking British Columbia with the other provinces within 10 years of its entry into Confederation. See B.C. Terms of Union, section 11.
4. See, for example, the essays in K. W. Studnicki-Gizbert, *Issues in Canadian Transport Policy* (Toronto: Macmillan of Canada, 1973) which omit any discussion of constitutional transport obligations as a significant factor in contemporary transport policy. See also C. Dalfen and L. Dunbar, "Transportation and Communications: The Constitution and the Canadian Economic Union," in *Case Studies in the Division of Powers*, Vol. 62, ed. Mark Krasnick (Royal Commission on the Economic Union and Development Prospects for Canada, Background Studies, 1985) which makes no mention of the federal obligations to provide transportation services.
5. See *Prince Edward Island (Minister of Transportation and Public Works) v. Canadian National Railway Co.* [1991] 1 F.C. 129.
6. See *Attorney General of British Columbia v. Attorney General of Canada* 42 B.C.L.R. (2d) 339 (1989).
7. See *Attorney General of British Columbia v. Attorney General of Canada* (Reasons for Judgment of the Court of Appeal for British Columbia, October 4, 1991, unreported).
8. It is important to distinguish constitutional transportation obligations from constitutional *jurisdiction* relating to transportation. Constitutional jurisdiction grants authority to deal with a particular matter, without specifying how that authority is to be exercised; a constitutional obligation imposes a duty to deal with a particular matter in a particular fashion. This study examines the transportation obligations imposed on the federal government under the Canadian Constitution.
9. P. B. Waite, *The Life and Times of Confederation* (Toronto: University of Toronto Press, 1967).
10. Glazebrook, p. 14.
11. It should be noted that the delegates from Prince Edward Island were opposed to the Intercolonial, regarding it as a cause for additional taxation of Island residents without any corresponding benefit. Thus the reference to "Maritime" support for the Intercolonial is limited to the provinces of New Brunswick and Nova Scotia.
12. G. P. Browne, ed., "Quebec Resolutions," *Documents on the Confederation of British North America* (Toronto: McClelland and Stewart, 1969), p. 165.
13. *Ibid.* p. 165.
14. The various drafts of the sections of the BNA Act setting out the commitment to build the Intercolonial are found in Browne, p. 278 and pp. 335–36.
15. Earlier drafts of the bill had contemplated specifying a date for completion of the railway; see Browne, p. 278 (3rd draft of bill, 2 February 1867).
16. S. Fleming, *The Intercolonial* (Montreal: Dawson Brothers, 1876), p. 76.

17. See section 146 of the BNA Act, 1867.
18. See section 2 of the *Manitoba Act, 1870*.
19. Rupert's Land and North-Western Territory Order, 1870 (U.K.), R.S.C. 1970, Appendix II, no. 9.
20. The text of the B.C. proposal was as follows:
 8. Inasmuch as no real union can subsist between this Colony and Canada without the speedy establishment of communications across the Rocky Mountains by Coach Road and Railway, the Dominion shall, within three years from the date of union, construct and open for traffic such Coach Road from some point on the line of the Main Trunk Road of this Colony to Fort Garry, of similar character to the said Main Trunk Road; and shall further engage to use all means in her power to complete such Railway communication at the earliest practicable date, and that Surveys to determine the proper line for such Railway shall be at once commenced; and that a sum of not less than One Million Dollars shall be expended in every year, from and after three years from the date of union, in actually constructing the initial sections of such Railway from the Seaboard of British Columbia to connect with the Railway system of Canada.

Papers in Connection with the Construction of the Canadian Pacific Railway (Carnarvon Papers), Vol. I, no. 2. (1890).
21. See Glazebrook, pp. 48–49.
22. Note, however, that the obligation in Term 11 was merely imposed on the “Government of the Dominion” rather than upon “Parliament,” as was the case with the Intercolonial. This indicates that the obligation in Term 11 does not bind Parliament directly in the sense that it does not oblige Parliament actually to enact legislation. Of course, Term 11 is still “constitutionally entrenched” in the sense that any legislation which Parliament does enact must not conflict with its provisions. Further, the courts will rely on Term 11 to strike down any legislation which is inconsistent with its terms. See discussion *infra*, pp. 36 and 56.
23. See, i.e. Glazebrook, p. 47.
24. *Debate on the Subject of Confederation with Canada*, Victoria, B.C. (G.P.O., 1870), p. 92.
25. *Carnarvon Papers*, Vol. 1, no. 2, clause 10.
26. See clause 7 (re San Francisco) and clause 10 (re coastal mail service) of the original B.C. proposals, reprinted in *Carnarvon Papers*, Vol. 1, no. 2, pp. 140–41.
27. *Debate on the Subject of Confederation with Canada*, Victoria, B.C., p. 78.
28. This issue was not raised in the B.C. debates on the terms; no notice is taken of the difference in the wording proposed for the service to San Francisco as opposed to the service to Olympia. Thus it is impossible to state whether the drafters of the B.C. proposal envisaged that the requirement to “supply” the ferry service to San Francisco would terminate at some point.

29. In addition to the obligations outlined above, the Canadian government also undertook to guarantee for 10 years the interest on a loan for the construction of a "first-class Graving Dock" at Esquimalt. Note that here the obligation was limited to the guarantee of interest; Canada did not have to construct or secure the construction of the Graving Dock. See Term 12 of the Terms of Union.
30. For a full account of P.E.I.'s opposition to the Quebec resolutions, see F. W. P. Bolger, *Prince Edward Island and Confederation* (Charlottetown: St. Dunstan's University Press, 1964), pp. 89–107.
31. See Bolger, p. 201.
32. For an account of the public reaction to the 1869 Canadian offer and the demand for construction of an Island railway as the price for entering Confederation see Bolger, pp. 202–13.
33. Bolger, p. 217.
34. *Ibid.*, p. 235.
35. The Terms of Union provided as follows: "That the railways under contract and in the course of construction for the Government of the Island, shall be the property of Canada."
36. The Island's debt stood at approximately \$3.7 million while the "debt allowance" amounted to over \$4.7 million.
37. The relevant term provided that: "Efficient Steam Service for the conveyance of mails and passengers, to be established and maintained between the Island and the mainland of the Dominion, Winter and Summer, thus placing the Island in continuous communication with the Intercolonial railway and the railway system of the Dominion."
38. Indeed, this point was noted during the debates in the P.E.I. Legislative Council on the proposed Terms of Union. One of the members objected to the fact that there was nothing in the terms requiring Canada to keep the railway in operation. The Premier replied that this was indeed the case, but that the inclusion of such a requirement was regarded as unnecessary. See *Debates and Proceedings of the Legislative Council of Prince Edward Island for the Session of 1873*, May 1, 1873, (Charlottetown: Queen's Printer), pp. 49–50.
39. See *Debates and Proceedings of the Legislative Council of Prince Edward Island for the Session of 1873*, p. 67.
40. This was added to the terms at the very end of the negotiations and was one of the final concessions from Canada which secured the support of the Island government. See Bolger, p. 279.
41. Canada also agreed to take a dredge boat that was under construction at a cost of up to \$22,000. However, there was no obligation on the part of Canada beyond the mere purchase of the boat.
42. S.C. 1867, c. 13.
43. See E. Forbes, *The Maritime Rights Movement: A Study in Canadian Regionalism* (Montreal: McGill-Queen's University Press, 1979).

44. The increased cost of construction was attributable to a variety of causes, not the least of which was incompetence and mismanagement by the Federal Board of Commissioners. See G. R. Stevens, *Canadian National Railways*, Vol. 1 (Toronto: Clark, Irwin, 1960), pp. 192–203.
45. See *An Act to amend the Act respecting the construction of the Intercolonial Railway*, S.C. 1874, c. 15.
46. See *An Act to incorporate Canadian National Railway Company*, S.C. 1919, c. 13.
47. The story is well known and will not be recounted here. See, e.g. D. Creighton, *John A. Macdonald: The Old Chieftain* (Toronto: Macmillan, 1955), and H. A. Innis, *A History of the Canadian Pacific Railway* (Toronto: University of Toronto Press, 1971).
48. The wording of Term 11 itself did not support this contention. Term 11 provided that the railway would connect the “Seaboard” of British Columbia with the railway system of Canada: there is no mention of a railway on Vancouver Island. Notwithstanding the wording of Term 11, the Macdonald Government had passed an Order in Council in 1873 fixing Esquimalt on Vancouver Island as the terminus of the Canadian Pacific Railway. It was apparently contemplated that a bridge would be constructed between the mainland and Vancouver Island near Nanaimo, and a line of railway would then be built on the Island from Nanaimo to Esquimalt. The plans for the bridge across Seymour Narrows never materialized.
49. See *Carnarvon Papers*, for the history of the negotiations between 1873 and 1883.
50. The Agreement dealt with a variety of matters in addition to the issue of the Island railway. For example, a change in the route of the CPR through the Rockies had resulted in the land grant in that area consisting of barren and mountainous country. The province agreed to provide compensation for the low value of this land by conveying a huge block of land (3.5 million acres) in the Peace River area. The Agreement also dealt with the terms and conditions respecting the sale of railway lands to the public.
51. Note that under the terms of the Settlement Agreement, the federal government itself did not undertake the obligation to construct the railway line. Its obligation was simply to transfer lands to the railway contractors, to pay a subsidy of \$750,000 and to take security from the contractors in order to ensure that the contractors would complete the line on schedule. See section (e) of the Settlement Agreement.
52. See *An Act relating to the Island Railway, the Graving Dock and Railway Lands of the Province*, S.B.C. 1884, c. 14. and *An Act respecting the Vancouver Island Railway, the Esquimalt Graving Dock, and certain Railway Lands of the Province of British Columbia*, granted to the Dominion, S.C. 1884, c. 6.
53. Section 16 of the contract provided:

The Canadian Pacific Railway, and all stations and station grounds, work shops, buildings, yards and other property, rolling stock and appurtenances required and used for the construction and working thereof, and the capital stock of the Company, shall be forever free from taxation by the Dominion, or by any Province hereafter to be established, or by any Municipal Corporation therein; and the lands of the Company, in the North-West Territories, until they are either sold or occupied, shall also be free from such taxation for 20 years after the grant thereof from the Crown.

54. See section 7. The obligation of the railway in this case was more clearly expressed than in the case of the Agreement with the Dunsmuir Syndicate regarding the Island railway. The obligation imposed on the CPR to operate the main line of railway "thereafter and forever" did not appear in the Agreement with the Dunsmuir group. The significance of this wording will be explored later in this paper in connection with the recent B.C. court decision on the operation of the Island railway.
55. See *An Act respecting the Canadian Pacific Railway*, S.C. 1881, c. 1.
56. See, e.g. *Esquimalt and Nanaimo Ry. v. A.G.B.C.* [1950] A.C. 87, dealing with the scope of the tax exemption for the lands owned by the CPR on Vancouver Island.
57. The British Columbia courts have recently held otherwise, concluding that certain statutory provisions dealing with the operation of the Vancouver Island Railway have become constitutionally entrenched. This case and its implications will be discussed later in section V of this paper.
58. See [1950] A.C. 87.
59. An exception arose in the case of Saskatchewan and Alberta in 1905. The federal statutes creating those provinces included the tax exemption granted to the CPR under the 1881 contract. (See section 24 of both the *Alberta Act* and the *Saskatchewan Act*, reproducing section 16 of the contract between the CPR and the Government of Canada). These federal statutes became constitutionally entrenched as part of the Constitution of Canada in 1982 and now have full constitutional force and effect. This means they can only be modified in accordance with the amending formula set out in part V of the *Constitution Act, 1982*.
60. See "Report of the Deputy Minister of Trade and Commerce," p. 39 in *The Thirty-fifth Annual Report of the Department of Trade and Commerce for the fiscal year ending March 31, 1927*.
61. *Ibid.*
62. The terms of the P.E.I. memorial are described in detail by Mr. Justice Cattanach in *The Queen in Right of the Province of Prince Edward Island v. The Queen in Right of Canada* [1976] 2 F.C. 712, p. 718.
63. S.C. 1901, c. 3, s.1.
64. See *The Prince Edward Island Subsidy Act, 1912*, S.C. 1912, c. 42, s. 2. This statute provided for an annual subsidy of \$100,000 to the Island, of which one half was attributable to the failure to provide efficient ferry service in accordance with the Terms of Union.
65. See *Report of the Commission of Inquiry into Newfoundland Transportation*, Vol. 1 (Sullivan Commission) (St. John's, 1978), pp. 30-31.
66. The Sullivan Commission stated that the railway had been a drain on the Newfoundland treasury of approximately \$750,000 annually ever since 1923.
67. Documents on Relations Between Canada and Newfoundland, no. 805.
68. Sullivan Commission Report, p. 34.
69. See *Re Residential Tenancies Act, 1979* [1981] 1 S.C.R. 714, pp. 721-23.

70. See *Reference Re Section 94(2) of the Motor Vehicle Act (B.C.)* [1985] 2 S.C.R. 486 [holding that the intentions of the drafters of the Charter of Rights, as manifested by statements made before a Parliamentary Committee, were entitled to minimal weight in judicial interpretation.]
71. Order in Council P.C. 1454, April 1, 1949.
72. Term 32(1): "Canada will maintain in accordance with the traffic offering a freight and passenger steamship service between North Sydney and Port aux Basques, which on completion of a motor highway between Corner Brook and Port aux Basques, will include suitable provision for the carriage of motor vehicles."
73. See Sullivan Commission Report, p. 34.
74. See *Attorney General of Newfoundland v. C.N.R.*, (1951) 67 C.R.T.C., p. 353.
75. See *The Queen (P.E.I.) v. The Queen (Can)* [1976] 2 F.C. 712 affd. [1978] 1 F.C. 533.
76. For a description of these impacts, see the judgement of the trial division of the Federal Court: [1976] 2 F.C. 712, p. 728.
77. [1976] 2 F.C. 712, p. 729. The trial judge also held that the federal government's obligation included provision for the conveyance of automobiles as well as passengers. The 1873 Terms of Union did not refer to automobiles, since they were non-existent at the time. But Cattanach J. applied section 10 of the *Interpretation Act*, R.S.C. 1970, c. I-23 to hold that the only sensible modern interpretation of the Terms of Union was one that included automobiles.
78. See [1976] 2 F.C. 712, p. 737.
79. *Ibid.*, p. 734.
80. [1978] 1 F.C. 533, p. 556. In the end, Jackett C.J. concluded that the Court could only issue a declaration that Prince Edward Island had a right to money compensation, as opposed to a judgement ordering the payment. But there was no practical difference between the issuance of a declaration as opposed to an actual judgement since in both cases the federal government would comply with the finding of the Court.
81. While Chief Justice Jackett did not express any final opinion on the point, he indicated that it was probably not open to individuals to seek compensation for the loss of ferry service. He held that the obligation was owed to the "Province," which he defined as "the mass of inhabitants of the geographical area whoever they may be from time to time." He indicated that this obligation to the Province did not give rise to rights in individuals or in groups of individuals. See *ibid.*, pp. 555-56 and footnote 30.
82. Le Dain J. made the point that the provincial claim should be limited to damages or costs imposed on the government directly. The province should not be permitted to obtain a judgement for the adverse effects which the province as a whole might have suffered. See *ibid.*, p. 589.
83. For a complete listing of the relevant statutes and orders, see "Schedule I" to the *Constitution Act, 1982*.
84. This follows from the holding in *Canada v. P.E.I.*

85. *Minister of Justice of Canada v. Borowski* (1981) 130 D.L.R. (3d) 588, p. 589 per Martland J. (S.C.C.).
86. [1976] 2 F.C. 712, p. 731 (per Cattanach J.).
87. See Order no. 1989-R-180.
88. See *P.E.I. v. CNR* [1991] 1 F.C., p. 129.
89. Iacobucci J.'s conclusion was as follows:

It is clear that Canada was to obtain property in the railway presumably because of Canada's assumption of the debts and liabilities of Prince Edward Island at the time of Union. Once it obtained the property, it was legally free to do what it wished with the railway as owner thereof. If an obligation to operate perpetually were intended, clear language to that effect would have been employed as was done, as Counsel for CN pointed out, in the 1883 B.C. railway settlement.
90. In *P.E.I. v. CNR*, Justice Iacobucci also makes the point that the obligation to maintain a ferry service only extends to passengers and mails, but not freight. This is the same approach adopted by Ledain J. in the Federal Court of Appeal in *Canada v. P.E.I.* On the other hand, Chief Justice Jaccett in the *Canada v. P.E.I.* case was of the view that automobiles *were included* within the constitutional obligation. This point does not seem of much practical significance, since the vessels providing the service all are equipped to transport automobiles and freight.
91. See generally the discussion of these developments in the decision of the Railway Transport Committee of the Canadian Transport Commission permitting the substitution of bus service for passenger trains: Pamphlet No. 14, 58 R.T.C., p. 359, July 3, 1968.
92. See opinion of D.H. Jones, Chairman, *ibid.*, p. 371.
93. *Ibid.*, p. 376.
94. See *Canadian National Railway Co. v. Board of Commissioners of Public Utilities* [1976] 2 S.C.R. 112.
95. See *Report of the Commission of Inquiry into Newfoundland Transportation*, Vol. 1 (1978), p. 233-34.
96. See "Remarks by the Honourable John Crosbie on the Newfoundland Transportation Initiative," June 20, 1988, p. 4. (Mimeographed.)
97. See "Memorandum of Understanding between the Government of Canada and the Government of the Province of Newfoundland," (MOU), June 20, 1988, paragraph 5. (Mimeographed.)
98. See *In the Matter of a complaint made by Atlantic Container Express Inc.* (National Transportation Agency of Canada, Decision No. 266-R-1991) May 22, 1991.
99. See *A.G. Nova Scotia v. A.G. Canada* (Nova Scotia Interdelegation) [1951] S.C.R. 31.
100. See *Reference Re. Canada Assistance Plan Act*, (S.C.C., August 15, 1991, unreported).

101. "Statement by the Honourable A. Brian Peckford, Premier of Newfoundland and Labrador on the Newfoundland Transportation Initiative," June 20, 1988, p. 7. (Mimeographed.) Obviously such a statement cannot itself alter any existing constitutional obligation. It does, however, give an indication of the legal opinion of the Government of the day with respect to the extent of the constitutional obligation.
102. All three members of the Agency were in agreement on this point; there was a dissenting opinion from one member of the board on the method to be used to calculate a Terms of Union rate. See *Atlantic Container Express Inc.*
103. *Ibid.*, p. 4.
104. See the attempt by the National Transportation Agency in its recent decision on the complaint by Atlantic Container Express to deal with this and other problems in establishing a Terms of Union rate.
105. In fact, as noted earlier, article 11 of the Terms of Union did not even mention the line on Vancouver Island; the obligation was to build a railway to the "Seaboard" of British Columbia.
106. 42 B.C.L.R. 339, p. 359 (1989).
107. *Ibid.*, p. 362.
108. The operators of the CPR promised to "thereafter and forever efficiently maintain, work and run the Canadian Pacific Railway." See S.C. 1881, c.1, section 7.
109. See Nova Scotia Interdelegation case, MOU between the Government of Canada and the Government of Newfoundland and Labrador.
110. As defined by section 52 of the *Constitution Act, 1982*.
111. See *Attorney General of British Columbia v. Attorney General of Canada* (Reasons for Judgment, B.C. Court of Appeal, October 4th, 1991, unreported).
112. Mr. Justice Esson, at trial, had found that the obligation to maintain service on the E. and N. was "perpetual"; this part of his Order was varied by the Court of Appeal.
113. See Reasons for Judgement, p. 41.
114. See, in particular, the Court's reasoning on pp. 44-45.
115. It is interesting to note that the Court of Appeal finds that the obligation is owed to the provincial government only. The implication is that private citizens are not entitled to rely upon these particular constitutional obligations. As noted previously, the obligations in Term 11 are part of the "Constitution of Canada" and are part of the "supreme law of Canada." It thus seems difficult to understand why the obligation can only be enforced by governments as opposed to private citizens. The Court of Appeal does not discuss the issue but merely notes in passing that the obligation could be varied with the consent of the B. C. government.
116. See Reasons for Judgement, p. 43.
117. See Reasons for Judgement, p. 45.

118. The "Canadian Pacific Railway" is defined in an 1874 statute as including the main line of railway, running from "a point near to and south of Lake Nipissing" westward to "some point in British Columbia on the Pacific Ocean." Further, the railway is to include two branch lines, the first constructed in Ontario, the second in Manitoba. See *An Act to Provide for the construction of the Canadian Pacific Railway*, S.C. 1874, ss. 2 and 3.
119. It is not known at this time whether the Attorney General of Canada intends to appeal. Leave to appeal would have to be granted by the Supreme Court of Canada.
120. See "Subsidy Agreement between the Government of Canada and the Government of the Province of British Columbia," April 18, 1977.
121. This is subject only to the single exception noted in paragraph (i) below.
122. All of the constitutional provisions identified here could be amended under section 43 of the *Constitution Act, 1982*, which requires the consent of the Parliament of Canada as well as any province to which the provision applies.
123. See *Reference Re Canada Assistance Plan Act*, (S.C.C. August 15, 1991, unreported).
124. The problem is potential only since there is no indication at this time of any dissatisfaction on the part of provinces with the enforceability of agreements relating to transportation.
125. Under the proposed section 95A, once an agreement had been ratified by the province and by the Parliament of Canada, it would have the force of law and it could not be amended without the consent of both parties. A similar procedure is contemplated by the recent constitutional proposals published by the Government of Canada on September 24, 1991. See *Shaping Canada's Future Together: Proposals*, (Ottawa: Supply and Services Canada, 1991), p. 34.

© Minister of Supply and Services Canada 1992

Extracts from this publication may be reproduced for individual use without permission provided the source is fully acknowledged. However, reproduction of this publication in whole or in part for purposes of resale or redistribution requires written permission from the Minister of Supply and Services, c/o Permissions Office, Canada Communications Group, Ottawa, Canada K1A 0S9.

550300016



JUN 9 1993

